STELAR-VISION: Self-Topology-Aware Efficient Learning for Aligned Reasoning in Vision

Chen Li
Carnegie Mellon University
chenli4@andrew.cmu.edu

Han Zhang
Carnegie Mellon University
hanz3@andrew.cmu.edu

Zhantao Yang
Carnegie Mellon University
zhantaoy@andrew.cmu.edu

Fangyi Chen
Carnegie Mellon University
fangyic@andrew.cmu.edu

Zihan Wang
Carnegie Mellon University
zihanw4@andrew.cmu.edu
fangyic@andrew.cmu.edu

Anudeepsekhar Bolimera
Carnegie Mellon University
abolimer@andrew.cmu.edu

Marios Saavides
Carnegie Mellon University
marioss@andrew.cmu.edu

Abstract

Vision-language models (VLMs) often rely on chain-of-thought (CoT) reasoning, resulting in verbose and suboptimal outputs on complex tasks. We introduce **STELAR-Vision**, a topology-aware training framework using **TopoAug** to generate diverse reasoning structures (Chain, Tree, Graph). Combined with supervised fine-tuning, reinforcement learning, and **Frugal Learning**, it improves both accuracy and efficiency—boosting Qwen2VL by 9.7%, surpassing Qwen2VL-72B by 7.3%, and outperforming Phi-4 and LLaMA-3.2 on five OOD benchmarks by up to 28.4% and 13.2%. We've released datasets, and code will be available.

1 Introduction

Recent advances in large language models (LLMs) have significantly improved reasoning capabilities, with models like GPT-o3 achieving strong performance on complex mathematical and scientific tasks. This progress has extended into the multimodal domain through vision-language models (VLMs) such as GPT-4o [OpenAI et al., 2024], GPT-4o-mini [OpenAI, 2024], and Qwen2.5-VL [Bai et al., 2025]. Despite the recent advances, there is still room for improvement in open-sourced VLMs when tackling complex vision-based reasoning tasks (e.g., math and science questions), and the path to enhance their abilities under an affordable training budget remains under-explored.

To address this, we begin by analyzing VLMs' reasoning behaviors and find that the popular models, both open-source and closed-source, tend to default to the chain-of-thought (CoT) [Wei et al., 2023] generation. However, our empirical analysis reveals that *different questions benefit from different reasoning topologies, such as Chain, Tree, or Graph structures* (Figure 4). The benefits of diverse reasoning topologies have yet to be well studied or effectively incorporated into existing training pipelines. Moreover, CoT often leads to verbose "overthinking", which increases the computational cost and makes real-time applications less viable. We find that there is a correlation between the topological reasoning structures and the output sequence length, thus providing insight into the overthinking problem created by the CoT reasoning.

39th Conference on Neural Information Processing Systems (NeurIPS 2025) Workshop: MATH-AI.

We propose Self-Topology-Aware-Efficient-Learning for Aligned Reasoning in Vision, **STELAR-Vision**, a topology-aware training framework using **TopoAug**, which generates and labels diverse reasoning structures. Models are post-trained via supervised fine-tuning (SFT) and reinforcement learning (RL) [Meng et al., 2024], and **Frugal Learning** is introduced to encourage concise, accurate outputs.

Our key contributions are: (1) We propose **STELAR-Vision**, a training framework that aligns diverse reasoning topologies such as chains, trees, and graphs with question characteristics. (2) We propose **TopoAug**, a synthetic pipeline that generates and labels structured reasoning paths for SFT and RL. (3) STELAR-Vision improves accuracy by **9.7%** over its base model and its larger variant Qwen2VL-72B-Instruct by **7.3%**. Frugal Learning reduces output length by **18.1%**.

2 Related Work

2.1 Topological Reasoning in Language and Vision Models

Chain-of-Thought (CoT) prompting [Wei et al., 2023] is a widely used reasoning strategy in LLMs and VLMs, guiding models to generate step-by-step solutions. However, its linear structure may not suit all tasks. To address this, Tree-of-Thought (ToT) [Yao et al., 2023] enables branching exploration, while Graph-of-Thought (GoT) [Besta et al., 2024] supports iterative and global reasoning. Both improve performance on complex tasks like TSP, algorithmic problem-solving, and multi-stage decision-making.

These methods, however, often rely on rule-based topology generations through sampling and are limited to language-only settings. In contrast, our framework automatically generates diverse topological structures and trains a VLM to adaptively select the optimal one per instance during decoding, enabling more flexible and generalizable reasoning.

2.2 Reinforcement Learning for LLM and VLM Reasoning

Reinforcement learning (RL) is a key technique for aligning LLMs and VLMs with desired behaviors in reasoning and preference modeling. Approaches like RLHF [Stiennon et al., 2022, Ouyang et al., 2022] and Constitutional AI [Bai et al., 2022] optimize models toward desired behaviors through preference feedback. Reward-based algorithms like PPO [Schulman et al., 2017], RPO [Yin et al., 2024], and GRPO [Shao et al., 2024] leverage explicit reward functions, while reward-free methods such as DPO [Rafailov et al., 2024], SimPO [Meng et al., 2024], and ORPO [Hong et al., 2024] achieve comparable results without reward modeling. These methods are widely applied to mathematical reasoning, long-horizon tasks, and instruction tuning.

In the vision-language domain, RL further enhances structured reasoning and safety-critical performance. Systems like VLM-RL [Huang et al., 2024], MedVLM-R1 [Pan et al., 2025], and RLVR [Chen et al., 2025] improve decision quality, medical safety, and out-of-distribution generalization.

Building on these insights, we show that combining topology-aware data generation with RL (e.g., SimPO) improves both accuracy and efficiency. Topological diversity expands the exploration space, increasing the likelihood of discovering stronger reasoning strategies during RL.

3 Method

In this section, we first construct topology-aware responses on two mathematical datasets. We then investigate the relationship between the topological reasoning and response accuracy. Finally, we present the topology-aware training framework shown in Figure 1.

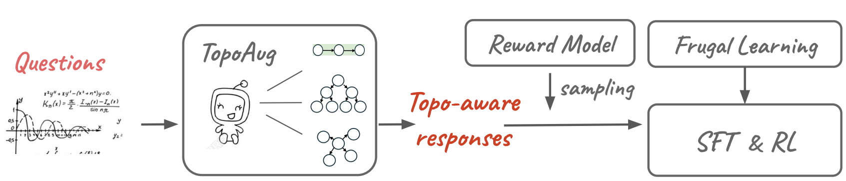


Figure 1: STELAR-Vision Framework.

3.1 Constructing Topology-Aware Responses

Data We use two math datasets: MATH-V Wang et al. [2024a] (3,040 visual problems) and VLM_S2H Park et al. [2025a] (7,000 logic puzzles), each pairing images with questions.

TopoAug: Generating Topology-Aware Responses We generate responses using topologies $T = \{Chain, Tree, Graph\}$, prompted via Qwen2-VL-7B-Instruct Wang et al. [2024b] and GPT-4o-Mini OpenAI [2024b] with extensive degrees of freedom in maximum depth, number of children, and number of neighbors. Please see the Figure 5 for detailed prompts.

Topology and Outcome Labels Each response r has an **Outcome Label** $\mathcal{H}_r \in \{0, 1\}$, and is assigned label 1 if correct and 0 otherwise. Each question-topology pair gets a **Topology Label** $\mathcal{F}_{q,t} = \frac{N_{\text{correct}}(q,t)}{N_{\text{total}}(q,t)}$ based on accuracy, where $N_{\text{correct}}(q, t)$ is the number of correct responses using topology t for question q, and $N_{\text{total}}(q, t)$ is the total number of responses generated using t.

Problem Difficulty Segmentation Problems are labeled Easy (>85th percentile), Hard (<15th), or Medium based on topology score distributions.

3.2 Analysis: Topological Reasoning Structures

We evaluate how reasoning topologies (*Chain*, *Tree*, *Graph*) affect performance by prompting VLMs under both default and guided settings. We compute a topology-wise *Win Rate* to assess the performance of each topological reasoning structure, which is defined below.

Win Rate We measure across the entire dataset and calculate the percentage of occurrence where a topology t is the best performing reasoning structure among the three topology types, computed as $\text{Win Rate}(t) = \sum_{q \in Q} \mathbb{1}_t(\text{argmax}_{t' \in T}\, \mathcal{F}_{q,t'})/N_Q$. Table 1 reports which topology wins most often.

	Chain	Tree	Graph
Win Rate	49%	28%	23%

Table 1: **Win Rates of reasoning topologies**. Combined Tree and Graph win rates exceed that of Chain.

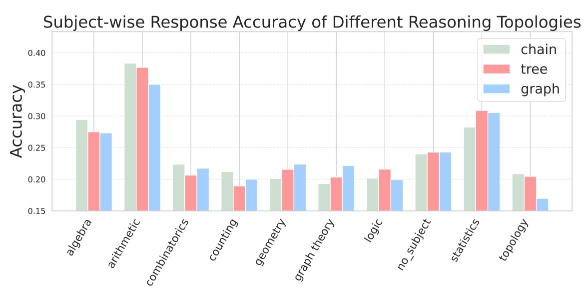


Figure 2: **Topology accuracy by subject:** Accuracy of $Chain$, $Tree$, and $Graph$ reasoning on MATH-V subjects. While $Chain$ is best overall, $Tree$ and $Graph$ excel in areas like graph theory and statistics.

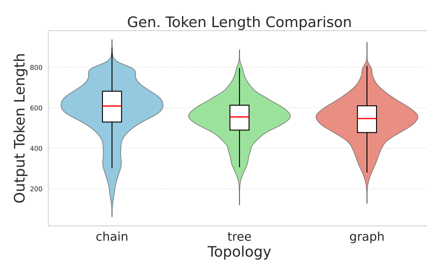


Figure 3: Token length distribution of $Chain$, $Tree$, and $Graph$ reasoning in TopoAug. Violin plots show median and interquartile range.

Table 1 shows $Chain$ performs best overall, but $Tree$ and $Graph$ win over half the time, signifying the great potential of dynamically choosing different topologies across different questions. Figure 2 shows subject-wise preferences, highlighting the benefit of matching topology to question type.

Topology-Wise Generation Length Figure 3 shows $Chain$ produces the longest responses with a right-skewed distribution that favors extended reasoning, while $Tree$ and $Graph$ are more concise.

3.3 STELAR-Vision Post-Training

Our findings motivate two assumptions: (1) training with diverse topologies helps models select optimal structures; (2) it encourage concise yet accurate outputs, enhancing inference efficiency. We implement this via a two-phase post-training pipeline: Supervised Fine-Tuning (SFT) and Reinforcement Learning (RL).

Model	In-Distribution Accuracy (%)			Out-of-Distribution Accuracy (%)				
	VLM_S2H	MATH-V	Overall	Geometry3K	We-Math	PolyMath	SciBench	LogicVista
GPT-4o OpenAI [2024a]	32.0	28.0	30.7	57.0	66.4	25.0	31.1	34.6
LLaVA-v1.6-Mistral-7B Liu et al. [2024]	26.0	8.0	18.0	20.6	26.0	9.2	3.4	18.5
Llama-3.2-11B-Vision-Instruct Grattafiori et al. [2024], Meta [2024]	22.0	10.0	18.0	35.0	37.8	22.2	10.7	24.8
MiniCPMv2.6-8B Yao et al. [2024]	1.5	13.0	18.7	45.0	50.2	14.4	8.5	20.7
Phi-4-multimodal-5.6B-instruct Abouelenin et al. [2025]	23.0	11.0	22.0	8.4	35.8	10.2	10.2	6.7
InternVL3-9B Zhu et al. [2025]	25.0	21.0	27.3	41.2	51.4	21.6	20.3	32.6
Qwen2VL-72B-Instruct Yang et al. [2024]	21.0	20.0	20.7	50.2	60.6	13.0	25.4	28.8
Qwen2VL-7B-Instruct Yang et al. [2024]	21.0	13.0	18.3	35.2	46.6	16.0	10.7	17.0
Chain-Only	25.0	21.0	23.7	31.4	42.2	17.2	10.7	25.4
STELAR-Vision-SFT	28.0	**24.0**	26.7	**44.4**	47.4	24.8	9.0	**33.3**
STELAR-Vision-RL-ONLY	24.0	23.0	23.7	32.8	39.0	**26.0**	**17.5**	23.9
STELAR-Vision	**31.0**	22.0	**28.0**	36.8	**51.0**	23.8	12.4	29.0

Table 2: **Quantitative Evaluation.** STELAR-Vision achieves strong gains on both ID and OOD tasks, outperforming its base by **9.7%**, Qwen2VL-72B by **7.3%**, and beating Phi-4 and LLaMA-3.2 by up to **36%** and **13.2%**, respectively. It also surpasses Chain-Only training by up to **13%**.

Phase 1: Supervised Fine-Tuning We fine-tune on TopoAug data mixed with OKVQA Marino et al. [2019], A-OKVQA Schwenk et al. [2022], and LLaVA150k Liu et al. [2023] (unaugmented for generalization). Data is filtered via: (1) balanced difficulty sampling, (2) correct responses only, and (3) rejection sampling with a 7B Outcome Reward Model. Training uses LoRA [Hu et al., 2021] with next-token prediction loss: $\mathcal{L}_{\text{NTP}} = -\sum_{t=1}^{T} \log P_\theta(y_t \mid y_{<t}, x)$.

Phase 2: Reinforcement Learning We initialize RL from the SFT checkpoint and apply SimPO [Meng et al., 2024] to prefer high-quality responses: $\mathcal{L}_{\text{SimPO}} = -\mathbb{E}[\log \sigma(\frac{\beta}{|y_w|} \log \pi(y_w) - \frac{\beta}{|y_l|} \log \pi(y_l) - \gamma)]$. We compare RL on TopoAug vs. Chain-only preferences (equal size), treating correct responses as preferred and removing topology prompts to enforce implicit structure learning.

Frugal Learning To improve efficiency, we introduce STELAR-Vision-Short, trained for concise yet accurate outputs. ***Variant 1 (†)*** prefers "short and correct" responses in both SFT and RL. ***Variant 2 (‡)*** further penalizes incorrect and overly long correct responses during RL to promote brevity.

Dataset	Subject	Is OOD?	Question Type	Sample Size
VLM_S2H	Math	✗	multiple-choice	200
MATH-V	Math	✗	free-form, multiple-choice	100
Geometry3K	Math	✓	multiple-choice	500
We-Math	Math	✓	multiple-choice	500
PolyMath	Math	✓	multiple-choice	500
SciBench	STEM	✓	free-form	177
LogicVista	Generic	✓	multiple-choice	448
Total				2425

Table 3: Summary of Evaluation Datasets

Dataset	**Sample**
MATH-V	85K
VLM_S2H	160K
OKVQA	18K
A-OKVQA	20K
LLava150k-inst	17K

Table 4: Total sample sizes of datasets used for training.

4 Experiments

4.1 Experimental Setup

Datasets We use 50K–60K samples from topology-augmented data (Section 3), plus 18k OKVQA, 36k A-OKVQA, and 17k LLaVA-150k for general multimodal tuning (Table 4).

Datasets and Models Evaluation spans MATH-V Wang et al. [2024a], VLM_S2H Park et al. [2025b], and five OOD benchmarks (Geometry3K Lu et al. [2021], We-Math Qiao et al. [2024], PolyMath Gupta et al. [2024], SciBench Wang et al. [2024c], LogicVista Xiao et al. [2024]), totaling 2,425 samples (Table 3). We use Qwen2VL-7B-Instruct Wang et al. [2024b] as our base model and compare with open/proprietary VLMs. Qwen2.5VL-7B Team [2024] is excluded due to instability.

Evaluation Metrics We report **Accuracy** and **Token Length** for performance and efficiency. Experiments run on 8×A100/H100(80GB) GPUs. Training takes ∼6 hours (SFT) and ∼9 hours (RL).

4.2 Overall Evaluation Results

We use STELAR-Vision to denote models trained with both phases of post-training, the `-SFT` suffix indicates models trained with supervised fine-tuning only, and `-RL-ONLY` indicates reinforcement learning directly from the base model without SFT. Table 2 shows results on the in-distribution datasets MATH-V and VLM_S2H as well as OOD datasets. STELAR-Vision achieves the highest in-distribution accuracy, significantly outperforming its base model Qwen2VL-7B-Instruct by **9.7%**, and surpassing larger model LLaMA-3.2-11B by **10.0%** and Qwen2VL-72B-Instruct by **7.3%**. While Chu et al. [2025] empirically finds SFT crucial before RL, our TopoAug-trained models outperform baselines with or without it, depending on the task. A deeper study is left for future work.

4.3 Ablation Studies

We perform ablation studies comparing our models to counterparts trained solely on chain-based reasoning data. Table 5 shows STELAR-Vision outperforms Chain-Only models by **4.3%** on in-distribution (ID) tasks and up to **8.8%** on out-of-distribution (OOD) datasets. This suggests our model learns to adaptively select reasoning topologies rather than relying on memorization.

Model	**w/ SFT**	**w/ RL**	**VLM_S2H**	**MATH-V**	**Overall**
Qwen2VL-7B-Instruct			21.0	13.0	18.3
Chain-Only-SFT	✓		18.5	19.0	18.7
Chain-Only-RL-ONLY		✓	23.5	15.0	20.7
Chain-Only	✓	✓	25.0	21.0	23.7
STELAR-Vision-SFT	✓		28.0	**24.0**	26.7
STELAR-Vision-RL-ONLY		✓	24.0	23.0	23.7
STELAR-Vision	✓	✓	**31.0**	22.0	**28.0**

Table 5: **Impact of TopoAug and Training.** On ID datasets, STELAR-Vision improves the best Chain-Only model from 25% to 31%, with a **4.3%** overall gain—demonstrating the value of topological augmentation.

4.4 Efficiency Gains from Frugal Learning

STELAR-Vision-Short† reduces token length by 101 (ID) and 24.5 (OOD) with minimal accuracy loss, while still outperforming Qwen2VL-7B-Instruct by 2.5%. Other variants like Chain-Only-Short† are less effective, showing the benefit of topology-aware Frugal Learning.

Model	**Accuracy (%)**	**Gen. Token Length**	
		ID	OOD
Qwen2VL-7B-Instruct	26.2	613.5	543.3
Chain-Only	28.7	878.4	742.6
Chain-Only-Short†	23.9	843.1	713.0
STELAR-Vision-SFT	26.7	604.4	483.3
STELAR-Vision	31.6	556.7	523.4
STELAR-Vision-Short†	28.7	**455.7**	**498.6**
STELAR-Vision-Short‡	21.9	538.9	555.9

Table 6: **Accuracy vs. token length:** STELAR-Vision improves accuracy with fewer tokens; Frugal Learning further enhances efficiency.

Model	Dataset	Tree	Graph	Chain
w/o STELAR-Vision	Overall	-	-	100.00
w/ STELAR-Vision	ID	14.3	9.7	76.0
	We-Math	63.0	7.4	29.6
	Geometry3K	96.4	3.0	0.6
	LogicVista	22.7	15.6	61.7
	PolyMATH	54.0	14.8	31.2
	SciBench	54.2	23.2	22.6

Table 7: **Test-time Topology Selection:** Distribution of reasoning topologies selected without prompting. ID denotes the in-distribution test split.

4.5 Why Our Method Works

STELAR-Vision improves ID and OOD tasks by enabling models to learn or select diverse reasoning topologies. As shown in Table 2, it outperforms Chain-only baselines, while Table 7 shows increased use of tree/graph structures, confirming the model adapts topology to problem type. Different datasets show distinct reasoning topology distributions; models adapt reasoning topologies to task complexity, using chains for simpler tasks and trees or graphs for harder ones. For models lacking this ability, our framework instills it via SFT and RL, though isolating effects is left for future work.

5 Conclusion and Limitation

We introduced STELAR-Vision, a topology-aware training framework that improves VLM reasoning by leveraging diverse structures. It outperforms its base by 9.7%, Qwen2VL-72B by 7.3%, and reduces output length by 18.1% with Frugal Learning—all while generalizing well across five OOD benchmarks. Despite strong results, STELAR-Vision relies on predefined topologies, and the link between problem structure and optimal reasoning remains underexplored. Future work will explore end-to-end topology induction and broader multimodal reasoning.

References

Abdelrahman Abouelenin, Atabak Ashfaq, Adam Atkinson, Hany Awadalla, Nguyen Bach, Jianmin Bao, Alon Benhaim, Martin Cai, Vishrav Chaudhary, Congcong Chen, et al. Phi-4-mini technical report: Compact yet powerful multimodal language models via mixture-of-loras. *arXiv preprint arXiv:2503.01743*, 2025.

Shuai Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Sibo Song, Kai Dang, Peng Wang, Shijie Wang, Jun Tang, Humen Zhong, Yuanzhi Zhu, Mingkun Yang, Zhaohai Li, Jianqiang Wan, Pengfei Wang, Wei Ding, Zheren Fu, Yiheng Xu, Jiabo Ye, Xi Zhang, Tianbao Xie, Zesen Cheng, Hang Zhang, Zhibo Yang, Haiyang Xu, and Junyang Lin. Qwen2.5-vl technical report, 2025. URL `https://arxiv.org/abs/2502.13923`.

Yuntao Bai, Saurav Kadavath, Sandipan Kundu, Amanda Askell, Jackson Kernion, Andy Jones, Anna Chen, Anna Goldie, Azalia Mirhoseini, Cameron McKinnon, Carol Chen, Catherine Olsson, Christopher Olah, Danny Hernandez, Dawn Drain, Deep Ganguli, Dustin Li, Eli Tran-Johnson, Ethan Perez, Jamie Kerr, Jared Mueller, Jeffrey Ladish, Joshua Landau, Kamal Ndousse, Kamile Lukosuite, Liane Lovitt, Michael Sellitto, Nelson Elhage, Nicholas Schiefer, Noemi Mercado, Nova DasSarma, Robert Lasenby, Robin Larson, Sam Ringer, Scott Johnston, Shauna Kravec, Sheer El Showk, Stanislav Fort, Tamera Lanham, Timothy Telleen-Lawton, Tom Conerly, Tom Henighan, Tristan Hume, Samuel R. Bowman, Zac Hatfield-Dodds, Ben Mann, Dario Amodei, Nicholas Joseph, Sam McCandlish, Tom Brown, and Jared Kaplan. Constitutional ai: Harmlessness from ai feedback, 2022. URL `https://arxiv.org/abs/2212.08073`.

Maciej Besta, Nils Blach, Ales Kubicek, Robert Gerstenberger, Michal Podstawski, Lukas Gianinazzi, Joanna Gajda, Tomasz Lehmann, Hubert Niewiadomski, Piotr Nyczyk, and Torsten Hoefler. Graph of thoughts: Solving elaborate problems with large language models. *Proceedings of the AAAI Conference on Artificial Intelligence*, 38(16):17682–17690, March 2024. ISSN 2159-5399. doi: 10.1609/aaai.v38i16.29720. URL `http://dx.doi.org/10.1609/aaai.v38i16.29720`.

Liang Chen, Lei Li, Haozhe Zhao, Yifan Song, Vinci, Lingpeng Kong, Qi Liu, and Baobao Chang. Rlvr in vision language models: Findings, questions and directions. *Notion Post*, Feb 2025. URL `https://deepagent.notion.site/rlvr-in-vlms`.

Tianzhe Chu, Yuexiang Zhai, Jihan Yang, Shengbang Tong, Saining Xie, Dale Schuurmans, Quoc V. Le, Sergey Levine, and Yi Ma. Sft memorizes, rl generalizes: A comparative study of foundation model post-training, 2025. URL `https://arxiv.org/abs/2501.17161`.

Aaron Grattafiori, Abhimanyu Dubey, Abhinav Jauhri, Abhinav Pandey, Abhishek Kadian, Ahmad Al-Dahle, Aiesha Letman, Akhil Mathur, Alan Schelten, Alex Vaughan, Amy Yang, Angela Fan, Anirudh Goyal, Anthony Hartshorn, Aobo Yang, Archi Mitra, Archie Sravankumar, Artem Korenev, Arthur Hinsvark, Arun Rao, Aston Zhang, Aurelien Rodriguez, Austen Gregerson, Ava Spataru, Baptiste Roziere, Bethany Biron, Binh Tang, Bobbie Chern, Charlotte Caucheteux, Chaya Nayak, Chloe Bi, Chris Marra, Chris McConnell, Christian Keller, Christophe Touret, Chunyang Wu, Corinne Wong, Cristian Canton Ferrer, Cyrus Nikolaidis, Damien Allonsius, Daniel Song, Danielle Pintz, Danny Livshits, Danny Wyatt, David Esiobu, Dhruv Choudhary, Dhruv Mahajan, Diego Garcia-Olano, Diego Perino, Dieuwke Hupkes, Egor Lakomkin, Ehab AlBadawy, Elina Lobanova, Emily Dinan, Eric Michael Smith, Filip Radenovic, Francisco Guzmán, Frank Zhang, Gabriel Synnaeve, Gabrielle Lee, Georgia Lewis Anderson, Govind Thattai, Graeme Nail, Gregoire Mialon, Guan Pang, Guillem Cucurell, Hailey Nguyen, Hannah Korevaar, Hu Xu, Hugo Touvron, Iliyan Zarov, Imanol Arrieta Ibarra, Isabel Kloumann, Ishan Misra, Ivan Evtimov, Jack Zhang, Jade Copet, Jaewon Lee, Jan Geffert, Jana Vranes, Jason Park, Jay Mahadeokar, Jeet Shah, Jelmer van der Linde, Jennifer Billock, Jenny Hong, Jenya Lee, Jeremy Fu, Jianfeng Chi, Jianyu Huang, Jiawen Liu, Jie Wang, Jiecao Yu, Joanna Bitton, Joe Spisak, Jongsoo Park, Joseph Rocca, Joshua Johnstun, Joshua Saxe, Junteng Jia, Kalyan Vasuden Alwala, Karthik Prasad, Kartikeya Upasani, Kate Plawiak, Ke Li, Kenneth Heafield, Kevin Stone, Khalid El-Arini, Krithika Iyer, Kshitiz Malik, Kuenley Chiu, Kunal Bhalla, Kushal Lakhotia, Lauren Rantala-Yeary, Laurens van der Maaten, Lawrence Chen, Liang Tan, Liz Jenkins, Louis Martin, Lovish Madaan, Lubo Malo, Lukas Blecher, Lukas Landzaat, Luke de Oliveira, Madeline Muzzi, Mahesh Pasupuleti, Mannat Singh, Manohar Paluri, Marcin Kardas, Maria Tsimpoukelli, Mathew Oldham, Mathieu Rita, Maya Pavlova, Melanie Kambadur, Mike Lewis, Min Si, Mitesh Kumar Singh, Mona Hassan, Naman Goyal, Narjes

Torabi, Nikolay Bashlykov, Nikolay Bogoychev, Niladri Chatterji, Ning Zhang, Olivier Duchenne, Onur Çelebi, Patrick Alrassy, Pengchuan Zhang, Pengwei Li, Petar Vasic, Peter Weng, Prajjwal Bhargava, Pratik Dubal, Praveen Krishnan, Punit Singh Koura, Puxin Xu, Qing He, Qingxiao Dong, Ragavan Srinivasan, Raj Ganapathy, Ramon Calderer, Ricardo Silveira Cabral, Robert Stojnic, Roberta Raileanu, Rohan Maheswari, Rohit Girdhar, Rohit Patel, Romain Sauvestre, Ronnie Polidoro, Roshan Sumbaly, Ross Taylor, Ruan Silva, Rui Hou, Rui Wang, Saghar Hosseini, Sahana Chennabasappa, Sanjay Singh, Sean Bell, Seohyun Sonia Kim, Sergey Edunov, Shaoliang Nie, Sharan Narang, Sharath Raparthy, Sheng Shen, Shengye Wan, Shruti Bhosale, Shun Zhang, Simon Vandenhende, Soumya Batra, Spencer Whitman, Sten Sootla, Stephane Collot, Suchin Gururangan, Sydney Borodinsky, Tamar Herman, Tara Fowler, Tarek Sheasha, Thomas Georgiou, Thomas Scialom, Tobias Speckbacher, Todor Mihaylov, Tong Xiao, Ujjwal Karn, Vedanuj Goswami, Vibhor Gupta, Vignesh Ramanathan, Viktor Kerkez, Vincent Gonguet, Virginie Do, Vish Vogeti, Vítor Albiero, Vladan Petrovic, Weiwei Chu, Wenhan Xiong, Wenyin Fu, Whitney Meers, Xavier Martinet, Xiaodong Wang, Xiaofang Wang, Xiaoqing Ellen Tan, Xide Xia, Xinfeng Xie, Xuchao Jia, Xuewei Wang, Yaelle Goldschlag, Yashesh Gaur, Yasmine Babaei, Yi Wen, Yiwen Song, Yuchen Zhang, Yue Li, Yuning Mao, Zacharie Delpierre Coudert, Zheng Yan, Zhengxing Chen, Zoe Papakipos, Aaditya Singh, Aayushi Srivastava, Abha Jain, Adam Kelsey, Adam Shajnfeld, Adithya Gangidi, Adolfo Victoria, Ahuva Goldstand, Ajay Menon, Ajay Sharma, Alex Boesenberg, Alexei Baevski, Allie Feinstein, Amanda Kallet, Amit Sangani, Amos Teo, Anam Yunus, Andrei Lupu, Andres Alvarado, Andrew Caples, Andrew Gu, Andrew Ho, Andrew Poulton, Andrew Ryan, Ankit Ramchandani, Annie Dong, Annie Franco, Anuj Goyal, Aparajita Saraf, Arkabandhu Chowdhury, Ashley Gabriel, Ashwin Bharambe, Assaf Eisenman, Azadeh Yazdan, Beau James, Ben Maurer, Benjamin Leonhardi, Bernie Huang, Beth Loyd, Beto De Paola, Bhargavi Paranjape, Bing Liu, Bo Wu, Boyu Ni, Braden Hancock, Bram Wasti, Brandon Spence, Brani Stojkovic, Brian Gamido, Britt Montalvo, Carl Parker, Carly Burton, Catalina Mejia, Ce Liu, Changhan Wang, Changkyu Kim, Chao Zhou, Chester Hu, Ching-Hsiang Chu, Chris Cai, Chris Tindal, Christoph Feichtenhofer, Cynthia Gao, Damon Civin, Dana Beaty, Daniel Kreymer, Daniel Li, David Adkins, David Xu, Davide Testuggine, Delia David, Devi Parikh, Diana Liskovich, Didem Foss, Dingkang Wang, Duc Le, Dustin Holland, Edward Dowling, Eissa Jamil, Elaine Montgomery, Eleonora Presani, Emily Hahn, Emily Wood, Eric-Tuan Le, Erik Brinkman, Esteban Arcaute, Evan Dunbar, Evan Smothers, Fei Sun, Felix Kreuk, Feng Tian, Filippos Kokkinos, Firat Ozgenel, Francesco Caggioni, Frank Kanayet, Frank Seide, Gabriela Medina Florez, Gabriella Schwarz, Gada Badeer, Georgia Swee, Gil Halpern, Grant Herman, Grigory Sizov, Guangyi, Zhang, Guna Lakshminarayanan, Hakan Inan, Hamid Shojanazeri, Han Zou, Hannah Wang, Hanwen Zha, Haroun Habeeb, Harrison Rudolph, Helen Suk, Henry Aspegren, Hunter Goldman, Hongyuan Zhan, Ibrahim Damlaj, Igor Molybog, Igor Tufanov, Ilias Leontiadis, Irina-Elena Veliche, Itai Gat, Jake Weissman, James Geboski, James Kohli, Janice Lam, Japhet Asher, Jean-Baptiste Gaya, Jeff Marcus, Jeff Tang, Jennifer Chan, Jenny Zhen, Jeremy Reizenstein, Jeremy Teboul, Jessica Zhong, Jian Jin, Jingyi Yang, Joe Cummings, Jon Carvill, Jon Shepard, Jonathan McPhie, Jonathan Torres, Josh Ginsburg, Junjie Wang, Kai Wu, Kam Hou U, Karan Saxena, Kartikay Khandelwal, Katayoun Zand, Kathy Matosich, Kaushik Veeraraghavan, Kelly Michelena, Keqian Li, Kiran Jagadeesh, Kun Huang, Kunal Chawla, Kyle Huang, Lailin Chen, Lakshya Garg, Lavender A, Leandro Silva, Lee Bell, Lei Zhang, Liangpeng Guo, Licheng Yu, Liron Moshkovich, Luca Wehrstedt, Madian Khabsa, Manav Avalani, Manish Bhatt, Martynas Mankus, Matan Hasson, Matthew Lennie, Matthias Reso, Maxim Groshev, Maxim Naumov, Maya Lathi, Meghan Keneally, Miao Liu, Michael L. Seltzer, Michal Valko, Michelle Restrepo, Mihir Patel, Mik Vyatskov, Mikayel Samvelyan, Mike Clark, Mike Macey, Mike Wang, Miquel Jubert Hermoso, Mo Metanat, Mohammad Rastegari, Munish Bansal, Nandhini Santhanam, Natascha Parks, Natasha White, Navyata Bawa, Nayan Singhal, Nick Egebo, Nicolas Usunier, Nikhil Mehta, Nikolay Pavlovich Laptev, Ning Dong, Norman Cheng, Oleg Chernoguz, Olivia Hart, Omkar Salpekar, Ozlem Kalinli, Parkin Kent, Parth Parekh, Paul Saab, Pavan Balaji, Pedro Rittner, Philip Bontrager, Pierre Roux, Piotr Dollar, Polina Zvyagina, Prashant Ratanchandani, Pritish Yuvraj, Qian Liang, Rachad Alao, Rachel Rodriguez, Rafi Ayub, Raghotham Murthy, Raghu Nayani, Rahul Mitra, Rangaprabhu Parthasarathy, Raymond Li, Rebekkah Hogan, Robin Battey, Rocky Wang, Russ Howes, Ruty Rinott, Sachin Mehta, Sachin Siby, Sai Jayesh Bondu, Samyak Datta, Sara Chugh, Sara Hunt, Sargun Dhillon, Sasha Sidorov, Satadru Pan, Saurabh Mahajan, Saurabh Verma, Seiji Yamamoto, Sharadh Ramaswamy, Shaun Lindsay, Shaun Lindsay, Sheng Feng, Shenghao Lin, Shengxin Cindy Zha, Shishir Patil, Shiva Shankar, Shuqiang Zhang, Shuqiang Zhang, Sinong Wang, Sneha Agarwal, Soji Sajuyigbe, Soumith Chintala, Stephanie Max, Stephen Chen, Steve Kehoe, Steve Satterfield, Sudarshan Govindaprasad, Sumit Gupta,

Summer Deng, Sungmin Cho, Sunny Virk, Suraj Subramanian, Sy Choudhury, Sydney Goldman, Tal Remez, Tamar Glaser, Tamara Best, Thilo Koehler, Thomas Robinson, Tianhe Li, Tianjun Zhang, Tim Matthews, Timothy Chou, Tzook Shaked, Varun Vontimitta, Victoria Ajayi, Victoria Montanez, Vijai Mohan, Vinay Satish Kumar, Vishal Mangla, Vlad Ionescu, Vlad Poenaru, Vlad Tiberiu Mihailescu, Vladimir Ivanov, Wei Li, Wenchen Wang, Wenwen Jiang, Wes Bouaziz, Will Constable, Xiaocheng Tang, Xiaojian Wu, Xiaolan Wang, Xilun Wu, Xinbo Gao, Yaniv Kleinman, Yanjun Chen, Ye Hu, Ye Jia, Ye Qi, Yenda Li, Yilin Zhang, Ying Zhang, Yossi Adi, Youngjin Nam, Yu, Wang, Yu Zhao, Yuchen Hao, Yundi Qian, Yunlu Li, Yuzi He, Zach Rait, Zachary DeVito, Zef Rosnbrick, Zhaoduo Wen, Zhenyu Yang, Zhiwei Zhao, and Zhiyu Ma. The llama 3 herd of models, 2024. URL `https://arxiv.org/abs/2407.21783`.

Himanshu Gupta, Shreyas Verma, Ujjwala Anantheswaran, Kevin Scaria, Mihir Parmar, Swaroop Mishra, and Chitta Baral. Polymath: A challenging multi-modal mathematical reasoning benchmark. *arXiv preprint arXiv:2410.14702*, 2024.

Jiwoo Hong, Noah Lee, and James Thorne. Orpo: Monolithic preference optimization without reference model, 2024. URL `https://arxiv.org/abs/2403.07691`.

Edward J. Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models, 2021. URL `https://arxiv.org/abs/2106.09685`.

Zilin Huang, Zihao Sheng, Yansong Qu, Junwei You, and Sikai Chen. Vlm-rl: A unified vision language models and reinforcement learning framework for safe autonomous driving, 2024. URL `https://arxiv.org/abs/2412.15544`.

Haotian Liu, Chunyuan Li, Qingyang Wu, and Yong Jae Lee. Visual instruction tuning, 2023. URL `https://arxiv.org/abs/2304.08485`.

Haotian Liu, Chunyuan Li, Yuheng Li, Bo Li, Yuanhan Zhang, Sheng Shen, and Yong Jae Lee. Llava-next: Improved reasoning, ocr, and world knowledge, January 2024. URL `https://llava-vl.github.io/blog/2024-01-30-llava-next/`.

Pan Lu, Ran Gong, Shibiao Jiang, Liang Qiu, Siyuan Huang, Xiaodan Liang, and Song-Chun Zhu. Inter-gps: Interpretable geometry problem solving with formal language and symbolic reasoning. *CoRR*, abs/2105.04165, 2021. URL `https://arxiv.org/abs/2105.04165`.

Kenneth Marino, Mohammad Rastegari, Ali Farhadi, and Roozbeh Mottaghi. Ok-vqa: A visual question answering benchmark requiring external knowledge, 2019. URL `https://arxiv.org/abs/1906.00067`.

Yu Meng, Mengzhou Xia, and Danqi Chen. Simpo: Simple preference optimization with a reference-free reward, 2024. URL `https://arxiv.org/abs/2405.14734`.

Meta. Llama-3.2-11b-vision-instruct, 2024. URL `https://huggingface.co/meta-llama/Llama-3.2-11B-Vision-Instruct`.

OpenAI. Gpt-4o technical report, 2024a. URL `https://openai.com/research/gpt-4o`. Accessed: Mar. 8, 2025.

OpenAI. Gpt-4o mini: Advancing cost-efficient intelligence, 2024b. URL `https://openai.com/index/gpt-4o-mini-advancing-cost-efficient-intelligence/`. Accessed: March 8, 2025.

OpenAI. Openai api: o4-mini model documentation. `https://platform.openai.com/docs/models/o4-mini`, 2024. Accessed May 2024.

OpenAI, :, Aaron Hurst, Adam Lerer, Adam P. Goucher, Adam Perelman, Aditya Ramesh, Aidan Clark, AJ Ostrow, Akila Welihinda, Alan Hayes, Alec Radford, Aleksander Mądry, Alex Baker-Whitcomb, Alex Beutel, Alex Borzunov, Alex Carney, Alex Chow, Alex Kirillov, Alex Nichol, Alex Paino, Alex Renzin, Alex Tachard Passos, Alexander Kirillov, Alexi Christakis, Alexis Conneau, Ali Kamali, Allan Jabri, Allison Moyer, Allison Tam, Amadou Crookes, Amin Tootoochian, Amin Tootoonchian, Ananya Kumar, Andrea Vallone, Andrej Karpathy, Andrew Braunstein,

Andrew Cann, Andrew Codispoti, Andrew Galu, Andrew Kondrich, Andrew Tulloch, Andrey Mishchenko, Angela Baek, Angela Jiang, Antoine Pelisse, Antonia Woodford, Anuj Gosalia, Arka Dhar, Ashley Pantuliano, Avi Nayak, Avital Oliver, Barret Zoph, Behrooz Ghorbani, Ben Leimberger, Ben Rossen, Ben Sokolowsky, Ben Wang, Benjamin Zweig, Beth Hoover, Blake Samic, Bob McGrew, Bobby Spero, Bogo Giertler, Bowen Cheng, Brad Lightcap, Brandon Walkin, Brendan Quinn, Brian Guarraci, Brian Hsu, Bright Kellogg, Brydon Eastman, Camillo Lugaresi, Carroll Wainwright, Cary Bassin, Cary Hudson, Casey Chu, Chad Nelson, Chak Li, Chan Jun Shern, Channing Conger, Charlotte Barette, Chelsea Voss, Chen Ding, Cheng Lu, Chong Zhang, Chris Beaumont, Chris Hallacy, Chris Koch, Christian Gibson, Christina Kim, Christine Choi, Christine McLeavey, Christopher Hesse, Claudia Fischer, Clemens Winter, Coley Czarnecki, Colin Jarvis, Colin Wei, Constantin Koumouzelis, Dane Sherburn, Daniel Kappler, Daniel Levin, Daniel Levy, David Carr, David Farhi, David Mely, David Robinson, David Sasaki, Denny Jin, Dev Valladares, Dimitris Tsipras, Doug Li, Duc Phong Nguyen, Duncan Findlay, Edede Oiwoh, Edmund Wong, Ehsan Asdar, Elizabeth Proehl, Elizabeth Yang, Eric Antonow, Eric Kramer, Eric Peterson, Eric Sigler, Eric Wallace, Eugene Brevdo, Evan Mays, Farzad Khorasani, Felipe Petroski Such, Filippo Raso, Francis Zhang, Fred von Lohmann, Freddie Sulit, Gabriel Goh, Gene Oden, Geoff Salmon, Giulio Starace, Greg Brockman, Hadi Salman, Haiming Bao, Haitang Hu, Hannah Wong, Haoyu Wang, Heather Schmidt, Heather Whitney, Heewoo Jun, Hendrik Kirchner, Henrique Ponde de Oliveira Pinto, Hongyu Ren, Huiwen Chang, Hyung Won Chung, Ian Kivlichan, Ian O'Connell, Ian O'Connell, Ian Osband, Ian Silber, Ian Sohl, Ibrahim Okuyucu, Ikai Lan, Ilya Kostrikov, Ilya Sutskever, Ingmar Kanitscheider, Ishaan Gulrajani, Jacob Coxon, Jacob Menick, Jakub Pachocki, James Aung, James Betker, James Crooks, James Lennon, Jamie Kiros, Jan Leike, Jane Park, Jason Kwon, Jason Phang, Jason Teplitz, Jason Wei, Jason Wolfe, Jay Chen, Jeff Harris, Jenia Varavva, Jessica Gan Lee, Jessica Shieh, Ji Lin, Jiahui Yu, Jiayi Weng, Jie Tang, Jieqi Yu, Joanne Jang, Joaquin Quinonero Candela, Joe Beutler, Joe Landers, Joel Parish, Johannes Heidecke, John Schulman, Jonathan Lachman, Jonathan McKay, Jonathan Uesato, Jonathan Ward, Jong Wook Kim, Joost Huizinga, Jordan Sitkin, Jos Kraaijeveld, Josh Gross, Josh Kaplan, Josh Snyder, Joshua Achiam, Joy Jiao, Joyce Lee, Juntang Zhuang, Justyn Harriman, Kai Fricke, Kai Hayashi, Karan Singhal, Katy Shi, Kavin Karthik, Kayla Wood, Kendra Rimbach, Kenny Hsu, Kenny Nguyen, Keren Gu-Lemberg, Kevin Button, Kevin Liu, Kiel Howe, Krithika Muthukumar, Kyle Luther, Lama Ahmad, Larry Kai, Lauren Itow, Lauren Workman, Leher Pathak, Leo Chen, Li Jing, Lia Guy, Liam Fedus, Liang Zhou, Lien Mamitsuka, Lilian Weng, Lindsay McCallum, Lindsey Held, Long Ouyang, Louis Feuvrier, Lu Zhang, Lukas Kondraciuk, Lukasz Kaiser, Luke Hewitt, Luke Metz, Lyric Doshi, Mada Aflak, Maddie Simens, Madelaine Boyd, Madeleine Thompson, Marat Dukhan, Mark Chen, Mark Gray, Mark Hudnall, Marvin Zhang, Marwan Aljubeh, Mateusz Litwin, Matthew Zeng, Max Johnson, Maya Shetty, Mayank Gupta, Meghan Shah, Mehmet Yatbaz, Meng Jia Yang, Mengchao Zhong, Mia Glaese, Mianna Chen, Michael Janner, Michael Lampe, Michael Petrov, Michael Wu, Michele Wang, Michelle Fradin, Michelle Pokrass, Miguel Castro, Miguel Oom Temudo de Castro, Mikhail Pavlov, Miles Brundage, Miles Wang, Minal Khan, Mira Murati, Mo Bavarian, Molly Lin, Murat Yesildal, Nacho Soto, Natalia Gimelshein, Natalie Cone, Natalie Staudacher, Natalie Summers, Natan LaFontaine, Neil Chowdhury, Nick Ryder, Nick Stathas, Nick Turley, Nik Tezak, Niko Felix, Nithanth Kudige, Nitish Keskar, Noah Deutsch, Noel Bundick, Nora Puckett, Ofir Nachum, Ola Okelola, Oleg Boiko, Oleg Murk, Oliver Jaffe, Olivia Watkins, Olivier Godement, Owen Campbell-Moore, Patrick Chao, Paul McMillan, Pavel Belov, Peng Su, Peter Bak, Peter Bakkum, Peter Deng, Peter Dolan, Peter Hoeschele, Peter Welinder, Phil Tillet, Philip Pronin, Philippe Tillet, Prafulla Dhariwal, Qiming Yuan, Rachel Dias, Rachel Lim, Rahul Arora, Rajan Troll, Randall Lin, Rapha Gontijo Lopes, Raul Puri, Reah Miyara, Reimar Leike, Renaud Gaubert, Reza Zamani, Ricky Wang, Rob Donnelly, Rob Honsby, Rocky Smith, Rohan Sahai, Rohit Ramchandani, Romain Huet, Rory Carmichael, Rowan Zellers, Roy Chen, Ruby Chen, Ruslan Nigmatullin, Ryan Cheu, Saachi Jain, Sam Altman, Sam Schoenholz, Sam Toizer, Samuel Miserendino, Sandhini Agarwal, Sara Culver, Scott Ethersmith, Scott Gray, Sean Grove, Sean Metzger, Shamez Hermani, Shantanu Jain, Shengjia Zhao, Sherwin Wu, Shino Jomoto, Shirong Wu, Shuaiqi, Xia, Sonia Phene, Spencer Papay, Srinivas Narayanan, Steve Coffey, Steve Lee, Stewart Hall, Suchir Balaji, Tal Broda, Tal Stramer, Tao Xu, Tarun Gogineni, Taya Christianson, Ted Sanders, Tejal Patwardhan, Thomas Cunninghman, Thomas Degry, Thomas Dimson, Thomas Raoux, Thomas Shadwell, Tianhao Zheng, Todd Underwood, Todor Markov, Toki Sherbakov, Tom Rubin, Tom Stasi, Tomer Kaftan, Tristan Heywood, Troy Peterson, Tyce Walters, Tyna Eloundou, Valerie Qi, Veit Moeller, Vinnie Monaco, Vishal Kuo, Vlad Fomenko, Wayne Chang, Weiyi Zheng, Wenda Zhou, Wesam Manassra,

Will Sheu, Wojciech Zaremba, Yash Patil, Yilei Qian, Yongjik Kim, Youlong Cheng, Yu Zhang, Yuchen He, Yuchen Zhang, Yujia Jin, Yunxing Dai, and Yury Malkov. Gpt-4o system card, 2024. URL https://arxiv.org/abs/2410.21276.

Long Ouyang, Jeff Wu, Xu Jiang, Diogo Almeida, Carroll L. Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, John Schulman, Jacob Hilton, Fraser Kelton, Luke Miller, Maddie Simens, Amanda Askell, Peter Welinder, Paul Christiano, Jan Leike, and Ryan Lowe. Training language models to follow instructions with human feedback, 2022. URL https://arxiv.org/abs/2203.02155.

Jiazhen Pan, Che Liu, Junde Wu, Fenglin Liu, Jiayuan Zhu, Hongwei Bran Li, Chen Chen, Cheng Ouyang, and Daniel Rueckert. Medvlm-r1: Incentivizing medical reasoning capability of vision-language models (vlms) via reinforcement learning, 2025. URL https://arxiv.org/abs/2502.19634.

Simon Park, Abhishek Panigrahi, Yun Cheng, Dingli Yu, Anirudh Goyal, and Sanjeev Arora. Generalizing from simple to hard visual reasoning: Can we mitigate modality imbalance in vlms? *arXiv preprint arXiv:2501.02669*, 2025a.

Simon Park, Abhishek Panigrahi, Yun Cheng, Dingli Yu, Anirudh Goyal, and Sanjeev Arora. Generalizing from simple to hard visual reasoning: Can we mitigate modality imbalance in vlms?, 2025b. URL https://arxiv.org/abs/2501.02669.

Runqi Qiao, Qiuna Tan, Guanting Dong, Minhui Wu, Chong Sun, Xiaoshuai Song, Zhuoma Gongque, Shanglin Lei, Zhe Wei, Miaoxuan Zhang, Runfeng Qiao, Yifan Zhang, Xiao Zong, Yida Xu, Muxi Diao, Zhimin Bao, Chen Li, and Honggang Zhang. We-math: Does your large multimodal model achieve human-like mathematical reasoning? *CoRR*, abs/2407.01284, 2024. doi: 10.48550/ARXIV.2407.01284. URL https://doi.org/10.48550/arXiv.2407.01284.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Stefano Ermon, Christopher D. Manning, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model, 2024. URL https://arxiv.org/abs/2305.18290.

John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. Proximal policy optimization algorithms, 2017. URL https://arxiv.org/abs/1707.06347.

Dustin Schwenk, Apoorv Khandelwal, Christopher Clark, Kenneth Marino, and Roozbeh Mottaghi. A-okvqa: A benchmark for visual question answering using world knowledge, 2022. URL https://arxiv.org/abs/2206.01718.

Zhihong Shao, Peiyi Wang, Qihao Zhu, Runxin Xu, Junxiao Song, Xiao Bi, Haowei Zhang, Mingchuan Zhang, Y. K. Li, Y. Wu, and Daya Guo. Deepseekmath: Pushing the limits of mathematical reasoning in open language models, 2024. URL https://arxiv.org/abs/2402.03300.

Nisan Stiennon, Long Ouyang, Jeff Wu, Daniel M. Ziegler, Ryan Lowe, Chelsea Voss, Alec Radford, Dario Amodei, and Paul Christiano. Learning to summarize from human feedback, 2022. URL https://arxiv.org/abs/2009.01325.

Qwen Team. Qwen2.5-vl-7b-instruct, 2024. URL https://huggingface.co/Qwen/Qwen2.5-VL-7B-Instruct. Accessed: Mar. 8, 2025.

Ke Wang, Junting Pan, Weikang Shi, Zimu Lu, Mingjie Zhan, and Hongsheng Li. Measuring multimodal mathematical reasoning with math-vision dataset. *arXiv preprint arXiv:2402.14804*, 2024a.

Peng Wang, Shuai Bai, Sinan Tan, Shijie Wang, Zhihao Fan, Jinze Bai, Keqin Chen, Xuejing Liu, Jialin Wang, Wenbin Ge, Yang Fan, Kai Dang, Mengfei Du, Xuancheng Ren, Rui Men, Dayiheng Liu, Chang Zhou, Jingren Zhou, and Junyang Lin. Qwen2-vl: Enhancing vision-language model's perception of the world at any resolution, 2024b. URL https://arxiv.org/abs/2409.12191.

Xiaoxuan Wang, Ziniu Hu, Pan Lu, Yanqiao Zhu, Jieyu Zhang, Satyen Subramaniam, Arjun R. Loomba, Shichang Zhang, Yizhou Sun, and Wei Wang. SciBench: Evaluating College-Level Scientific Problem-Solving Abilities of Large Language Models. In *Proceedings of the Forty-First International Conference on Machine Learning*, 2024c.

Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten Bosma, Brian Ichter, Fei Xia, Ed Chi, Quoc Le, and Denny Zhou. Chain-of-thought prompting elicits reasoning in large language models, 2023. URL https://arxiv.org/abs/2201.11903.

Yijia Xiao, Edward Sun, Tianyu Liu, and Wei Wang. Logicvista: Multimodal llm logical reasoning benchmark in visual contexts, 2024. URL https://arxiv.org/abs/2407.04973.

An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, Jianxin Yang, Jin Xu, Jingren Zhou, Jinze Bai, Jinzheng He, Junyang Lin, Kai Dang, Keming Lu, Keqin Chen, Kexin Yang, Mei Li, Mingfeng Xue, Na Ni, Pei Zhang, Peng Wang, Ru Peng, Rui Men, Ruize Gao, Runji Lin, Shijie Wang, Shuai Bai, Sinan Tan, Tianhang Zhu, Tianhao Li, Tianyu Liu, Wenbin Ge, Xiaodong Deng, Xiaohuan Zhou, Xingzhang Ren, Xinyu Zhang, Xipin Wei, Xuancheng Ren, Xuejing Liu, Yang Fan, Yang Yao, Yichang Zhang, Yu Wan, Yunfei Chu, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, Zhifang Guo, and Zhihao Fan. Qwen2 technical report, 2024. URL https://arxiv.org/abs/2407.10671.

Shunyu Yao, Dian Yu, Jeffrey Zhao, Izhak Shafran, Thomas L. Griffiths, Yuan Cao, and Karthik Narasimhan. Tree of thoughts: Deliberate problem solving with large language models, 2023. URL https://arxiv.org/abs/2305.10601.

Yuan Yao, Tianyu Yu, Ao Zhang, Chongyi Wang, Junbo Cui, Hongji Zhu, Tianchi Cai, Haoyu Li, Weilin Zhao, Zhihui He, et al. Minicpm-v: A gpt-4v level mllm on your phone. *arXiv preprint arXiv:2408.01800*, 2024.

Yueqin Yin, Zhendong Wang, Yi Gu, Hai Huang, Weizhu Chen, and Mingyuan Zhou. Relative preference optimization: Enhancing llm alignment through contrasting responses across identical and diverse prompts, 2024. URL https://arxiv.org/abs/2402.10958.

Jinguo Zhu, Weiyun Wang, Zhe Chen, Zhaoyang Liu, Shenglong Ye, Lixin Gu, Hao Tian, Yuchen Duan, Weijie Su, Jie Shao, Zhangwei Gao, Erfei Cui, Xuehui Wang, Yue Cao, Yangzhou Liu, Xingguang Wei, Hongjie Zhang, Haomin Wang, Weiye Xu, Hao Li, Jiahao Wang, Nianchen Deng, Songze Li, Yinan He, Tan Jiang, Jiapeng Luo, Yi Wang, Conghui He, Botian Shi, Xingcheng Zhang, Wenqi Shao, Junjun He, Yingtong Xiong, Wenwen Qu, Peng Sun, Penglong Jiao, Han Lv, Lijun Wu, Kaipeng Zhang, Huipeng Deng, Jiaye Ge, Kai Chen, Limin Wang, Min Dou, Lewei Lu, Xizhou Zhu, Tong Lu, Dahua Lin, Yu Qiao, Jifeng Dai, and Wenhai Wang. Internvl3: Exploring advanced training and test-time recipes for open-source multimodal models, 2025. URL https://arxiv.org/abs/2504.10479.

A Technical Appendices and Supplementary Material

A.1 Limitations of Chain-of-Thoughts

While Chain-of-Thoughts (CoT) has been a popular reasoning method, we identify its pain points and leverage dynamic reasoning topology to resolve the problem. We demonstrate examples of such limitations and our solution in Figure 4

A.2 Example Prompts for TopoAug

We show example questions and prompts used to generate TopoAug answers in Figure 5.

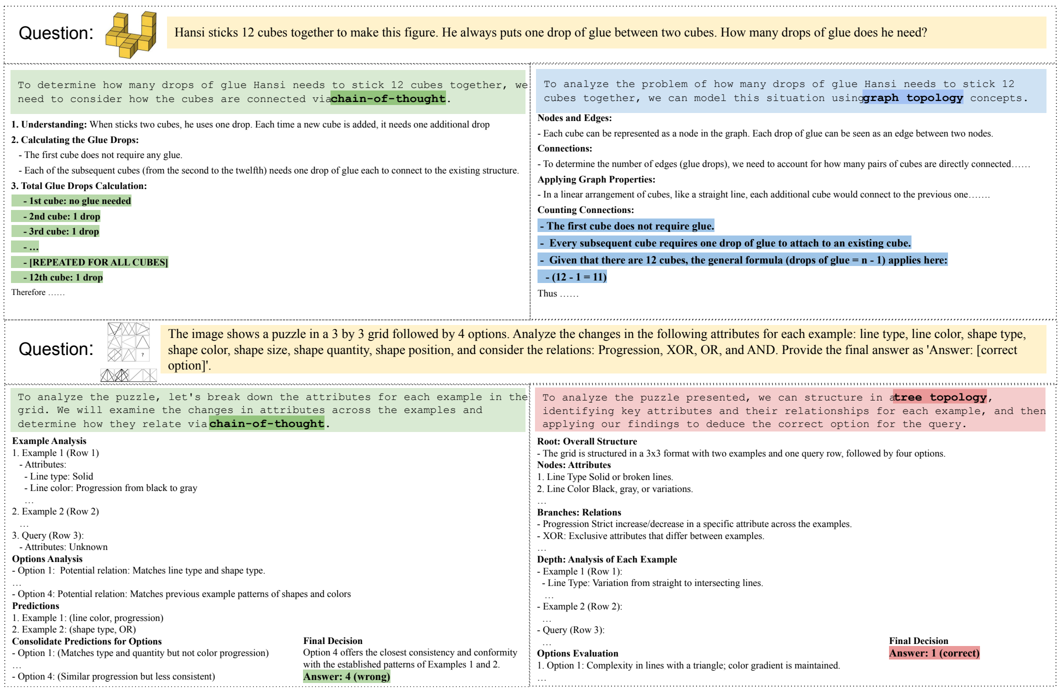


Figure 4: **Limitations of the Popular Chain-of-Thought Reasoning Structures.** The widely adopted Chain-of-Thought (CoT) reasoning paradigm (in green) often results in unnecessarily verbose reasoning processes, as demonstrated in the first example. Under CoT reasoning, the model redundantly counts each cube, whereas with $Graph$ topology (in blue), it quickly identifies the key point of the question. In the bottom-row example, CoT reasoning begins with a detailed examination of each subplot but ultimately arrives at an incorrect answer. In contrast, $Tree$ topology (in red) initiates reasoning with a high-level overview before delving into specific features. In both scenarios, CoT-style reasoning proves suboptimal.

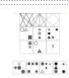
Original Question: 

In quadrilateral ABCD, angle ABC and angle ADC are both 90 degrees. The sides AD and DC are equal in length,. Additionally, the combined length of sides AB and BC is 20 centimeters. Given this information, what is the area of quadrilateral ABCD, measured in square centimeters?

[Prompt of Chain]:
In quadrilateral ABCD, angle ABC and angle ADC are both 90 degrees. The sides AD and DC are equal in length,. Additionally, the combined length of sides AB and BC is 20 centimeters. Given this information, what is the area of quadrilateral ABCD, measured in square centimeters? Please answer the question by first providing your reasoning of **chain topology**.

[Prompt of Tree]:
In quadrilateral ABCD, angle ABC and angle ADC are both 90 degrees. The sides AD and DC are equal in length,. Additionally, the combined length of sides AB and BC is 20 centimeters. Given this information, what is the area of quadrilateral ABCD, measured in square centimeters? Please answer the question by first providing your reasoning of **tree topology**.

[Prompt of Graph]:
In quadrilateral ABCD, angle ABC and angle ADC are both 90 degrees. The sides AD and DC are equal in length,. Additionally, the combined length of sides AB and BC is 20 centimeters. Given this information, what is the area of quadrilateral ABCD, measured in square centimeters? Please answer the question by first providing your reasoning of **graph topology**.

Original Question: 

The image shows a puzzle in a 3 by 3 grid followed by 4 options. The puzzle consists of 2 examples (row 1 and 2), a query (row 3), and four options. Each example contains three images following a relation along certain attribute, and this relation is consistent across all examples. The query contains two images. Analyze the changes in the following attributes for each example: line type, line color, shape type, shape color, shape size, shape quantity, shape position, and consider the relations: Progression, XOR, OR, and AND. Progression requires the value of a certain attribute to strictly increase or decrease, but not necessarily by a fixed amount. Please provide your predictions in the format 'Example i: (attribute, relation)' for each example and similarly for options. Provide the final answer as 'Answer: [correct option]'.

[Prompt of Chain]:
The image shows a puzzle in a 3 by 3 grid followed by 4 options. The puzzle consists of 2 examples (row 1 and 2), a query (row 3), and four options. Each example contains three images following a relation along certain attribute, and this relation is consistent across all examples. The query contains two images. Analyze the changes in the following attributes for each example: line type, line color, shape type, shape color, shape size, shape quantity, shape position, and consider the relations: Progression, XOR, OR, and AND. Progression requires the value of a certain attribute to strictly increase or decrease, but not necessarily by a fixed amount. Please provide your predictions in the format 'Example i: (attribute, relation)' for each example and similarly for options. Please answer the question by first providing your reasoning of **chain topology**, and then provide the final answer as 'Answer: [correct option]'.

[Prompt of Tree]:
The image shows a puzzle in a 3 by 3 grid followed by 4 options. The puzzle consists of 2 examples (row 1 and 2), a query (row 3), and four options. Each example contains three images following a relation along certain attribute, and this relation is consistent across all examples. The query contains two images. Analyze the changes in the following attributes for each example: line type, line color, shape type, shape color, shape size, shape quantity, shape position, and consider the relations: Progression, XOR, OR, and AND. Progression requires the value of a certain attribute to strictly increase or decrease, but not necessarily by a fixed amount. Please provide your predictions in the format 'Example i: (attribute, relation)' for each example and similarly for options. Please answer the question by first providing your reasoning of **tree topology**, and then provide the final answer as 'Answer: [correct option]'.

[Prompt of Graph]:
The image shows a puzzle in a 3 by 3 grid followed by 4 options. The puzzle consists of 2 examples (row 1 and 2), a query (row 3), and four options. Each example contains three images following a relation along certain attribute, and this relation is consistent across all examples. The query contains two images. Analyze the changes in the following attributes for each example: line type, line color, shape type, shape color, shape size, shape quantity, shape position, and consider the relations: Progression, XOR, OR, and AND. Progression requires the value of a certain attribute to strictly increase or decrease, but not necessarily by a fixed amount. Please provide your predictions in the format 'Example i: (attribute, relation)' for each example and similarly for options. Please answer the question by first providing your reasoning of **graph topology**, and then provide the final answer as 'Answer: [correct option]'.

Figure 5: Prompts used in TopoAug to generate training samples for the MathVision (MATH-V) and VLM_S2H datasets. For each question, we prompt the generation models to produce 10 to 20 responses across three distinct reasoning topologies—Chain, Tree, and Graph—with extensive flexibility in maximum depth, number of children, and number of neighbors.

A.3 Experimental Details

Model Architecture We used Qwen2-VL as the base model, where we keep the architecture not changed.

Training Hyperparameters We show detailed training hyperparameters we used in Table 8

	Hyperparameter	**Value**
Precision and Initialization	Precision	`bfloat16`
	Gradient checkpointing	Enabled
	FlashAttention-2	Enabled
Optimizer and Schedule	Optimizer	`AdamW`
	Learning rate	$\eta = 1.0 \times 10^{-5}$
	Weight decay	Not explicitly specified
	Scheduler	Cosine decay
	Warmup	10%
	Gradient accumulation steps	16
LoRA and PEFT	LoRA rank	16
	LoRA alpha	32
	`use_peft`	`True`
SimPO Parameters	α	2.5
	γ	1.38

Table 8: Summary of hyperparameters.

NeurIPS Paper Checklist

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Our main claims made in the abstract and introduction are accurate. Our main claims are further summarized in Figure 1 and Table 2, and illustrated with details in Section 2 and Section 3.

   Guidelines:

   - The answer NA means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A No or NA answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

   Justification: We have discussed the limitations of our work in Section 4.

   Guidelines:

   - The answer NA means that the paper has no limitation while the answer No means that the paper has limitations, but those are not discussed in the paper.
   - The authors are encouraged to create a separate "Limitations" section in their paper.
   - The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
   - The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
   - The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
   - The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
   - If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
   - While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

Answer: [NA]

Justification: This is not a theoretical paper.

Guidelines:

- The answer NA means that the paper does not include theoretical results.
- All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
- All assumptions should be clearly stated or referenced in the statement of any theorems.
- The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
- Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
- Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: We provided our experiments' settings in Section 2, 3 and Appendix A.3.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - If the paper includes experiments, a No answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
   - If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
   - Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
   - While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
     (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
     (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
     (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
     (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

Answer: [No]

Justification: While our codebase is not yet ready for public release at the time of submission, we provide detailed implementation descriptions to support reproducibility.

Guidelines:

- The answer NA means that paper does not include experiments requiring code.
- Please see the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- While we encourage the release of code and data, we understand that this might not be possible, so "No" is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
- The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://nips.cc/public/guides/CodeSubmissionPolicy`) for more details.
- The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
- The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
- At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).
- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer, etc.) necessary to understand the results?

   Answer: [Yes]

   Justification: We provided our experiments' settings in Section 2, 3 and Appendix A.3.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: Our paper primarily studies Large Vision Language Models with reasoning, therefore making it computationally costly and prohibited to run multiple times for each experiment for error bars calculation. However, we have shown that our performances consistently outperform the baseline method in Table 2 and 5 in multiple settings.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The authors should answer "Yes" if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.

- The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
- The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
- The assumptions made should be given (e.g., Normally distributed errors).
- It should be clear whether the error bar is the standard deviation or the standard error of the mean.
- It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
- For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g. negative error rates).
- If error bars are reported in tables or plots, The authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: It's detailed in Section 3.

   Guidelines:

   - The answer NA means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.
   - The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
   - The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We are convinced that we comply with NeurIPS Code of Ethics

   Guidelines:

   - The answer NA means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer No, they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [NA]

    Justification: Our work focuses on small-scale models and datasets using standard training protocols. It does not pose broader societal impact beyond advancing our understanding of specific aspects of deep learning.

    Guidelines:

    - The answer NA means that there is no societal impact of the work performed.

- If the authors answer NA or No, they should explain why their work has no societal impact or why the paper does not address societal impact.
- Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
- The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
- The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
- If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pretrained language models, image generators, or scraped datasets)?

    Answer: [NA]

    Justification: Our paper does not have such risks.

    Guidelines:

    - The answer NA means that the paper poses no such risks.
    - Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
    - Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
    - We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: We've credited and cited the references and codebases appropriately in the paper.

    Guidelines:

    - The answer NA means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.

- If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
- For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
- If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [NA]

    Justification: We don't release any new assets.

    Guidelines:

    - The answer NA means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [NA]

    Justification: Our research does not involve human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
    - According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

    Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

    Answer: [NA]

    Justification: Our paper does not involve crowdsourcing nor research with human subjects.

    Guidelines:

    - The answer NA means that the paper does not involve crowdsourcing nor research with human subjects.
    - Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.

- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

    Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does not impact the core methodology, scientific rigorousness, or originality of the research, declaration is not required.

    Answer: [Yes]

    Justification: Our paper uses LLMs and VLMs to generate reasoning data as part of our main method, and we provide more details and information of this generation and usage in Section 2 and 3. Our paper also uses the LLM for writing, editing, or formatting purposes and does not impact the scientific rigorousness, or originality of the research.

    Guidelines:

    - The answer NA means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
    - Please refer to our LLM policy (`https://neurips.cc/Conferences/2025/LLM`) for what should or should not be described.