KAT-V1: Kwai-AutoThink Technical Report

arxiv.org/abs/2507.08297

https://huggingface.co/Kwaipilot/KAT-V1-40B

Zizheng Zhan*†, Ken Deng*, Huaixi Tang*, Wen Xiang*, Kun Wu*, Weihao Li, Wenqiang Zhu, Jingxuan Xu, Lecheng Huang, Zongxian Feng, Shaojie Wang, Shangpeng Yan, Jiaheng Liu, Zhongyuan Peng, Zuchen Gao, Haoyang Huang, Ziqi Zhan, Yanan Wu, Yuanxing Zhang, Jian Yang, Guang Chen, Haotian Zhang, Bin Chen, Bing Yu

Kwaipilot Team

{zhanzizheng, dengken, zhanghaotian}@kuaishou.com

简介

为了解决llm在推理任务中存在的过度思考(overthinking)问题，本文提出KAT-V1 (Kwaipilot-AutoThink），一个能够自动决定是否进行推理的开源40B llm。KAT-V1 经历了三阶段训练：1）将Qwen2.5-32B upscaling (复制layer)到40B，然后构建Long/Short CoT训练集，结合Multi-Token Prediction(MTP)进行知识蒸馏，教会模型在外部trigger指令(<think_on>/<think_off>)下是否执行推理；2）sft，在训练数据中引入DeepSeek生成的判断是否要推理的judge信息，让模型学习根据query内容自主判断是否需要显式推理；3）基于Step-SRPO强化学习算法的RLVR训练。

背景

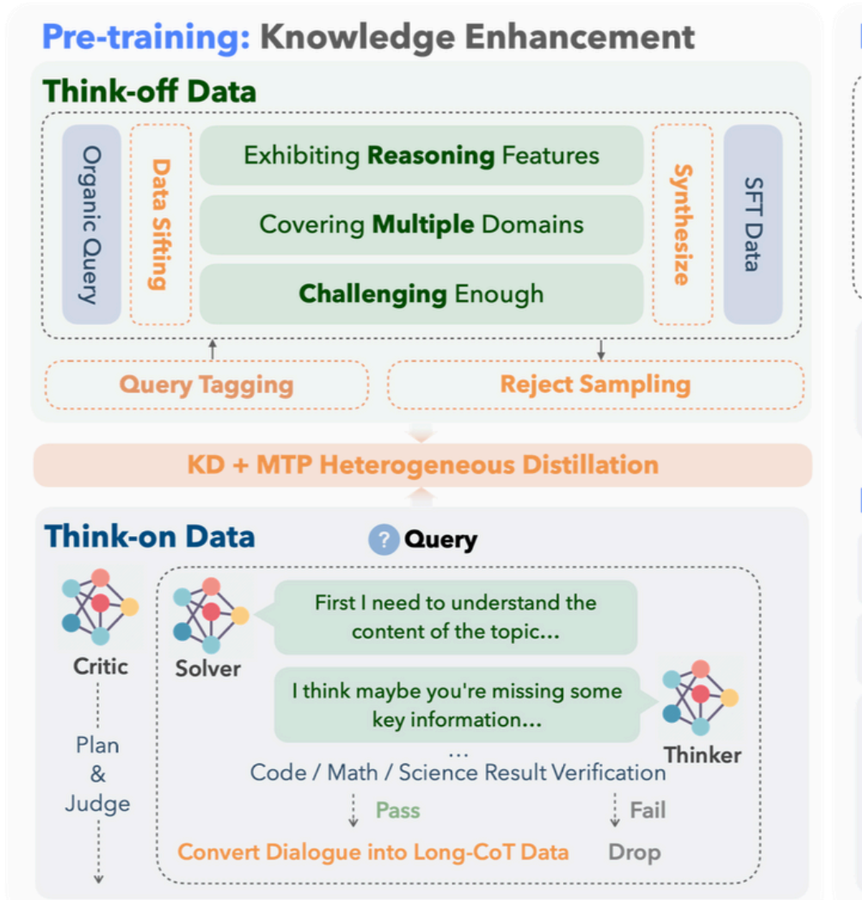
注意：KAT-V1的40B是从Qwen2.5-32B upscale得到的

虽然reasoning llm在一些复杂query上取得了相当好的效果，但是真的没必要对所有query都进行推理，比如“Apple的中文是什么？”这就不需要推理了吧。这就引出了一个问题：过度思考(overthinking)，即llm对对简单query生成不必要地的罗里吧嗦的冗长解释。如何让llm能根据query自动切换是否进行推理呢？我们前面读过几个相关工作了，今天看下快手KAT-V1模型给出的方案。

KAT-V1三阶段训练

- upscaling + Multi-Token Prediction (MTP) + 知识蒸馏
- cold start via sft
- Step-SRPO for RLVR

注意：论文说模型是两阶段训练的，
这里的三阶段是我根据自己的理解整理的，仅供参考



部分实验结果

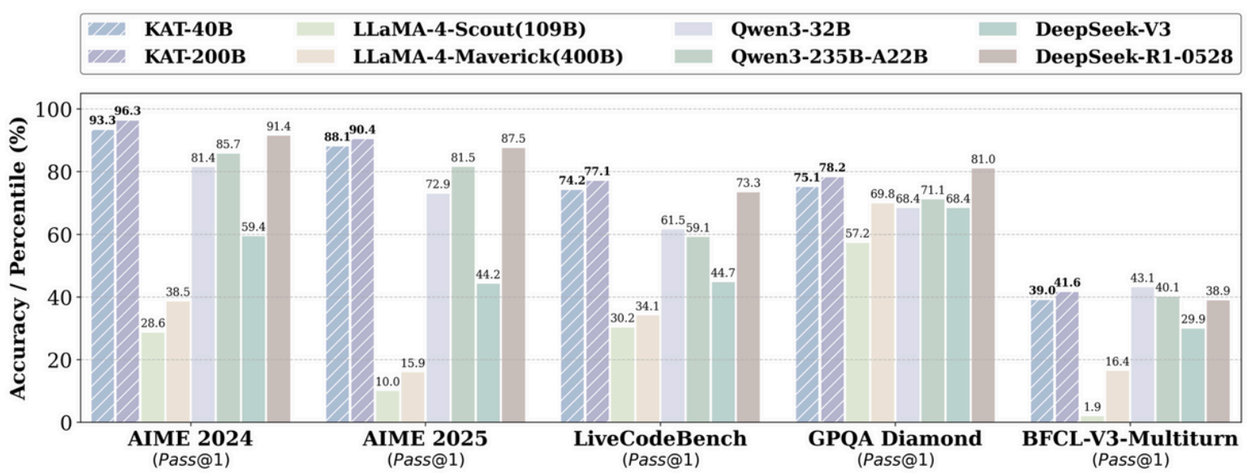


Figure 1 | Performance of Kwaipilot-AutoThink on various benchmarks.

思考

不论本文用了几个阶段的训练，结合之前读过的论文来看，目前让llm决定是否进行推理的做法，本质上还是在long/short CoT混合数据上训练。
由于本文写的不是特别详细，当然已经比不少“技术报告”细节多了，所以疑惑不少，最大的就是为什么要对Qwen2.5-32B做upsaling啊？也没看到消融实验对比对32B进行相同训练的效果要比40B差。还是很期待后续Step-SRPO论文的