

2024.11.28 - 2024.12.4

- [PoT](#) PoT的prompt都没怎么写，还得用CodeX模型，感觉不是理想。
- [PAL](#) 目前的实验基于PAL，PAL可以在GPT上使用。
- [Adversarial Math Word Problem Generation](#) 修改问题中的数值，甚至效果也会比较好。
- [Exploring the Compositional Deficiency of Large Language Models in Mathematical Reasoning](#) math trap问题，主要是math trap的分类。
- [RL on Incorrect Synthetic Data Scales the Efficiency of LLM Math Reasoning by Eight-Fold](#)
- [AI-ASSISTED GENERATION OF DIFFICULT MATH QUESTIONS](#) 这篇论文方法和我之前提出的第二个想法有一点点相似之处，就是从问题中提取知识点，这篇论文中要求必须使用其中两个知识点，再大模型合成数据集，人类检验并且完善问题。
- 我们如果想要生成很难的问题，应该只能使用专家检验了，我想。
- 如果必须是循环多次的计算问题，还算是数学问题吗？并不考察数学能力。