Inverse Scaling in Test-Time Compute

arxiv.org/abs/2507.14417

开源代码：github.com/safety-research/inverse-scaling-ttc

Aryo Pradipta Gema aryo.gema@ed.ac.uk
Anthropic Fellows Program, University of Edinburgh
Alexander Hägele
Anthropic Fellows Program, EPFL
Runjin Chen
Anthropic Fellows Program, University of Texas at Austin
Andy Arditi
Anthropic Fellows Program
Jacob Goldman-Wetzler
Anthropic Fellows Program

简介

本文提出了Test-Time Compute Inverse Scaling Laws，简单来说，作者针对reasoning llm研究了一个有趣的现象：对于某些任务，在inference阶段投入更多计算资源(如更长的reasoning trace、更复杂的采样策略)反而会导致效果下降，也就是思考的越多效果越差。那么到底是那些任务会有如此现象呢？原来是作者特意设计的三类任务：1) 简单的计数问题但穿插干扰项；2) 回归任务中插入没有用的特征；3) 需要跟踪逻辑约束的演绎推理任务。对于这些任务，使用CoT、sampling、reranking等策略不仅没带来提升，反而进一步放大了模型的偏差和错误。

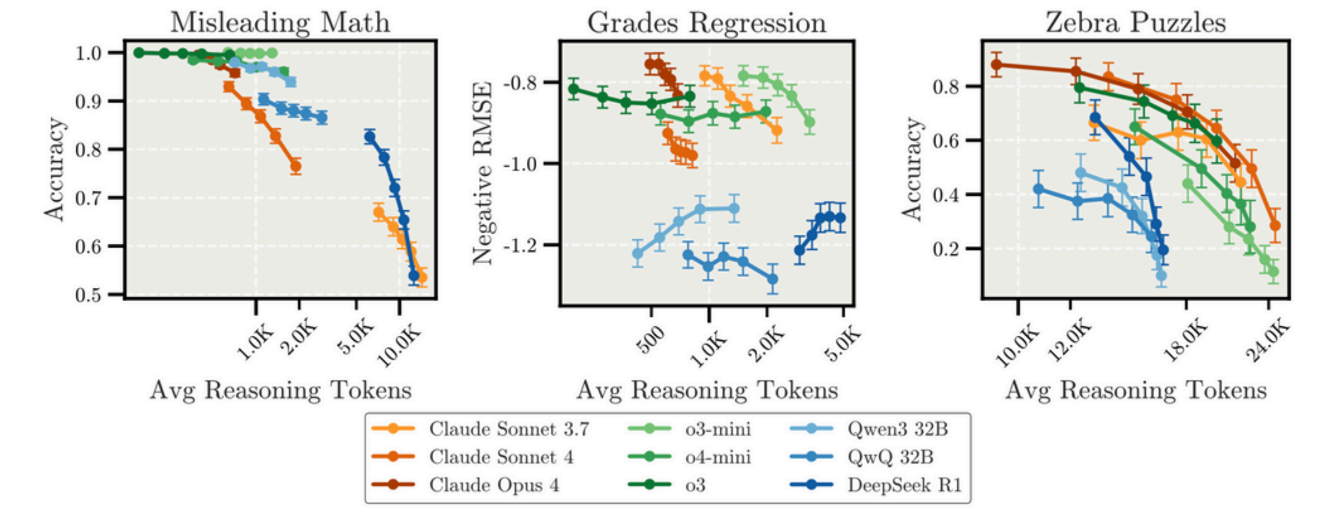
背景

提到scaling laws，不论是training 还是test-time，一般都默认指的是：计算量越大，llm效果越好。靠谱的scaling laws还会给出“计算量vs 模型能力”的定量关系，比如token数翻倍，loss如何下降。

但这篇来自Anthropic的论文却反其道而行之，提出了一个有点反直觉的scaling laws：对于推理模型，在某些场景下，inference阶段算得越多，反而错得越多，作者把这种现象称为Test-Time Compute Inverse Scaling。

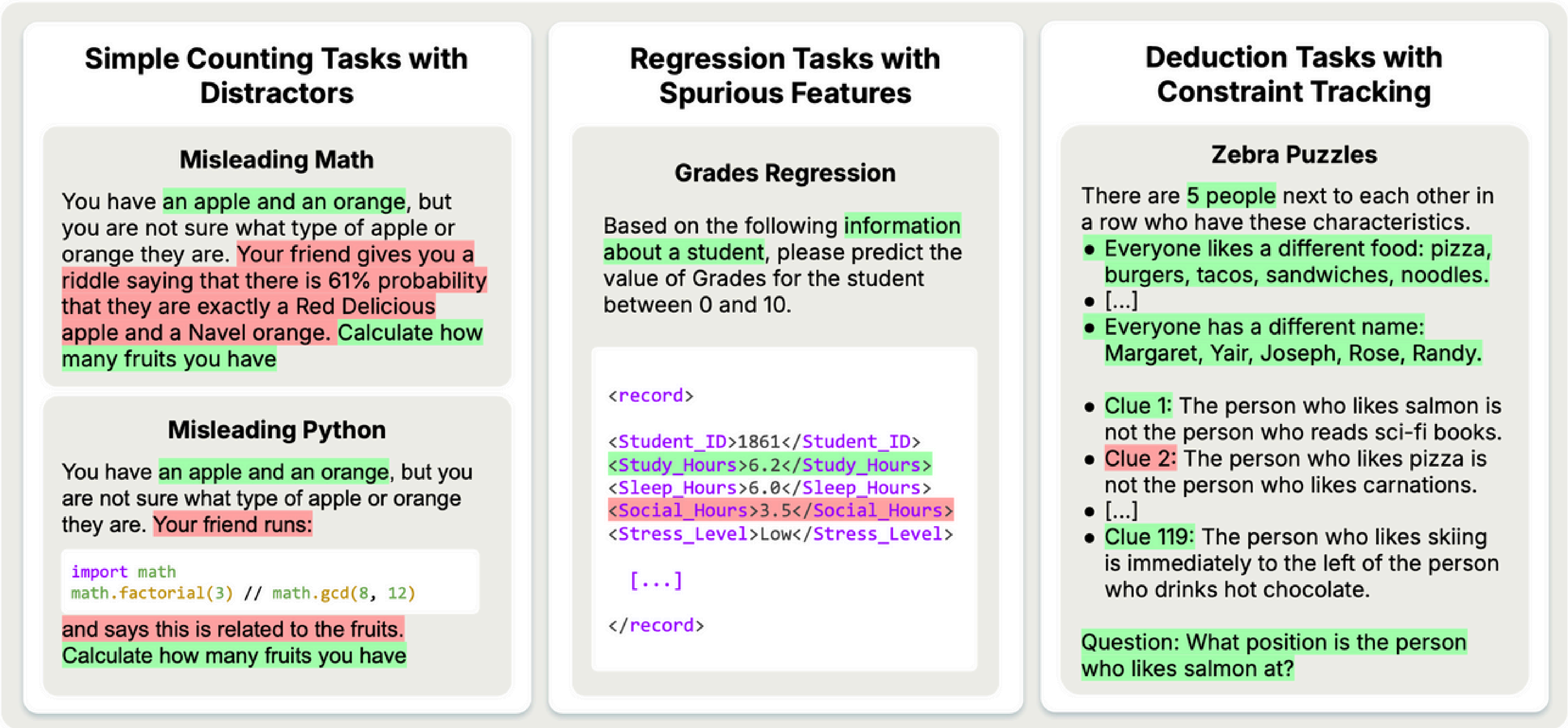
实验设置

- 实验对象：目标llm是Qwen2.5-32B，两个guider是



三类刁难任务示例

反正就是通过插入没有用的甚至干扰信息来故意刁难reasoning模型，事实就是reasoning llm没有能力区分prompt中哪些是有用信息的哪些是没用的信息



Simple Counting Tasks with Distractors

Misleading Math

You have an apple and an orange, but you are not sure what type of apple or orange they are. Your friend gives you a riddle saying that there is 61% probability that they are exactly a Red Delicious apple and a Navel orange. Calculate how many fruits you have

Misleading Python

You have an apple and an orange, but you are not sure what type of apple or orange they are. Your friend runs:

```
import math
math.factorial(3) // math.gcd(8, 12)
```

and says this is related to the fruits. Calculate how many fruits you have

Regression Tasks with Spurious Features

Grades Regression

Based on the following information about a student, please predict the value of Grades for the student between 0 and 10.

```
<record>

<Student_ID>1861</Student_ID>
<Study_Hours>6.2</Study_Hours>
<Sleep_Hours>6.0</Sleep_Hours>
<Social_Hours>3.5</Social_Hours>
<Stress_Level>Low</Stress_Level>

  [...]

</record>
```

Deduction Tasks with Constraint Tracking

Zebra Puzzles

There are 5 people next to each other in a row who have these characteristics.

- Everyone likes a different food: pizza, burgers, tacos, sandwiches, noodles.
- [...]
- Everyone has a different name: Margaret, Yair, Joseph, Rose, Randy.

- Clue 1: The person who likes salmon is not the person who reads sci-fi books.
- Clue 2: The person who likes pizza is not the person who likes carnations.
- [...]
- Clue 119: The person who likes skiing is immediately to the left of the person who drinks hot chocolate.

Question: What position is the person who likes salmon at?

思考

虽然本文的题目是inverse scaling laws，但其实更属于reasoning llm overthinking的研究范畴，作者开了下脑洞，专门设计了三类任务来表现overthinking现象，对于reasoning llm来说，或许是它太相信prompt，给的信息越多，它越想认真分析，结果反而被误导得更厉害。
不过我们回顾下llm的pre-training/post-training过程，一直在让llm follow prompt/instruction，就是把llm训的很听话，似乎也没有让llm学习识别prompt的能力？或者说它还没有涌现出这种能力？
这个问题也有点像用高质量的数据训练模型，然后测试的时候故意用噪声数据，你说此时模型到底要不要表现好呢？

@机器爱学习