

RAG + Prompt engineering  
提升 deepseek-r1-distill-qwen-14b  
解决电动力学问题的能力

第3组汇报展示

寇皓 黄逸涵 应嘉禾 刘储闻

# 初步尝试-微调

- 试图微调大预言模型，提高其解答数理问题的水平
  - Deepseek-R1模型开源，在解决数理问题上有很大优势
  - 对比发现：普通大语言模型不能很好地完成推理任务，选用推理模型
- > 选用 deepseek-r1-distill-qwen-14b 模型（后简称 R1 模型）

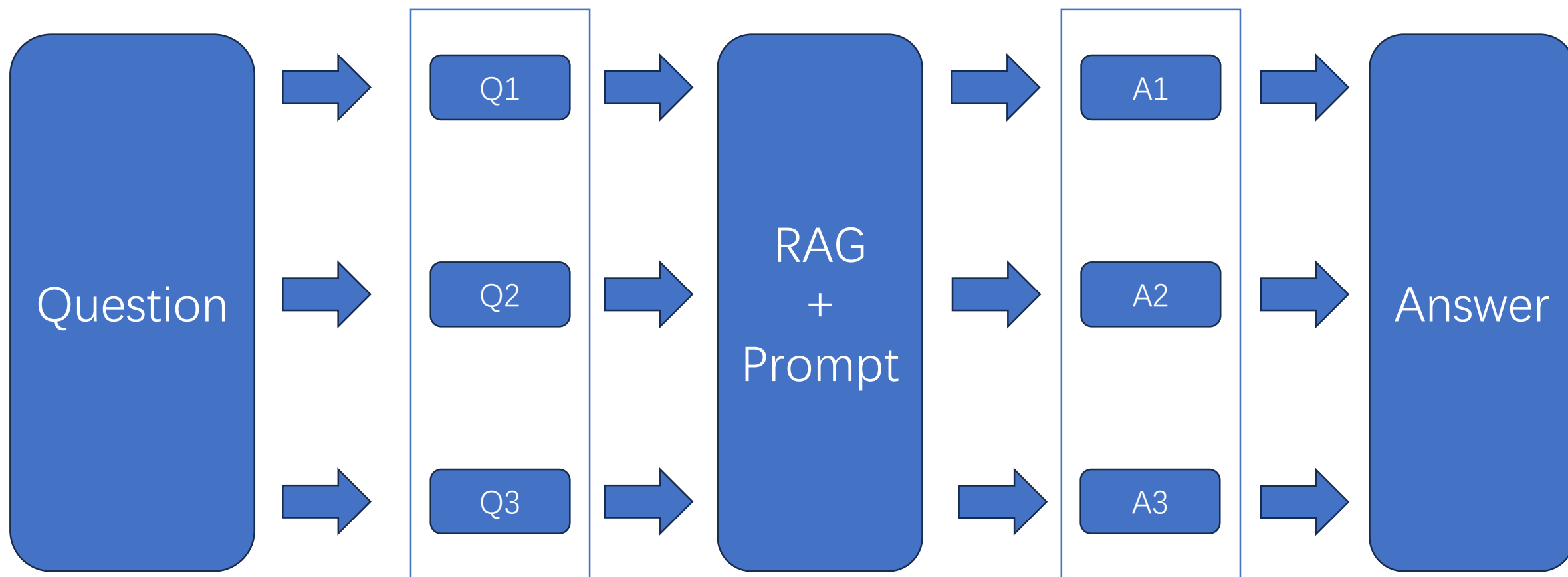
# R1 模型

- 得到结果不够准确
- 复杂问题的检索和解答耗时过长
- 思考长度过长，容易反复陷入自我怀疑

# 初步思路

- 使用 R1 模型对问题进行拆分，分步解决
- 让 R1 验证答案正确性并做出修正
- 通过 prompt engineering 引导模型解决问题
- 构建 RAG（检索增强生成），提供相关资料参考
  - RAG 模型能够把语料量化为向量，从海量数据中高效提取相关内容
  - R1 模型主动生成关键词，在 RAG 资料库中进行检索
- 用 Wolfram 处理 R1 无法计算的相关公式

迭代



# 初步结果

- R1 思考问题时，会直接把题目做完，并且会自觉分步思考问题

-> 问题分步，逐个给 R1 思考不适用

-> 使用 prompt engineering 引导模型更精确地分步作答

- 生成时间限制

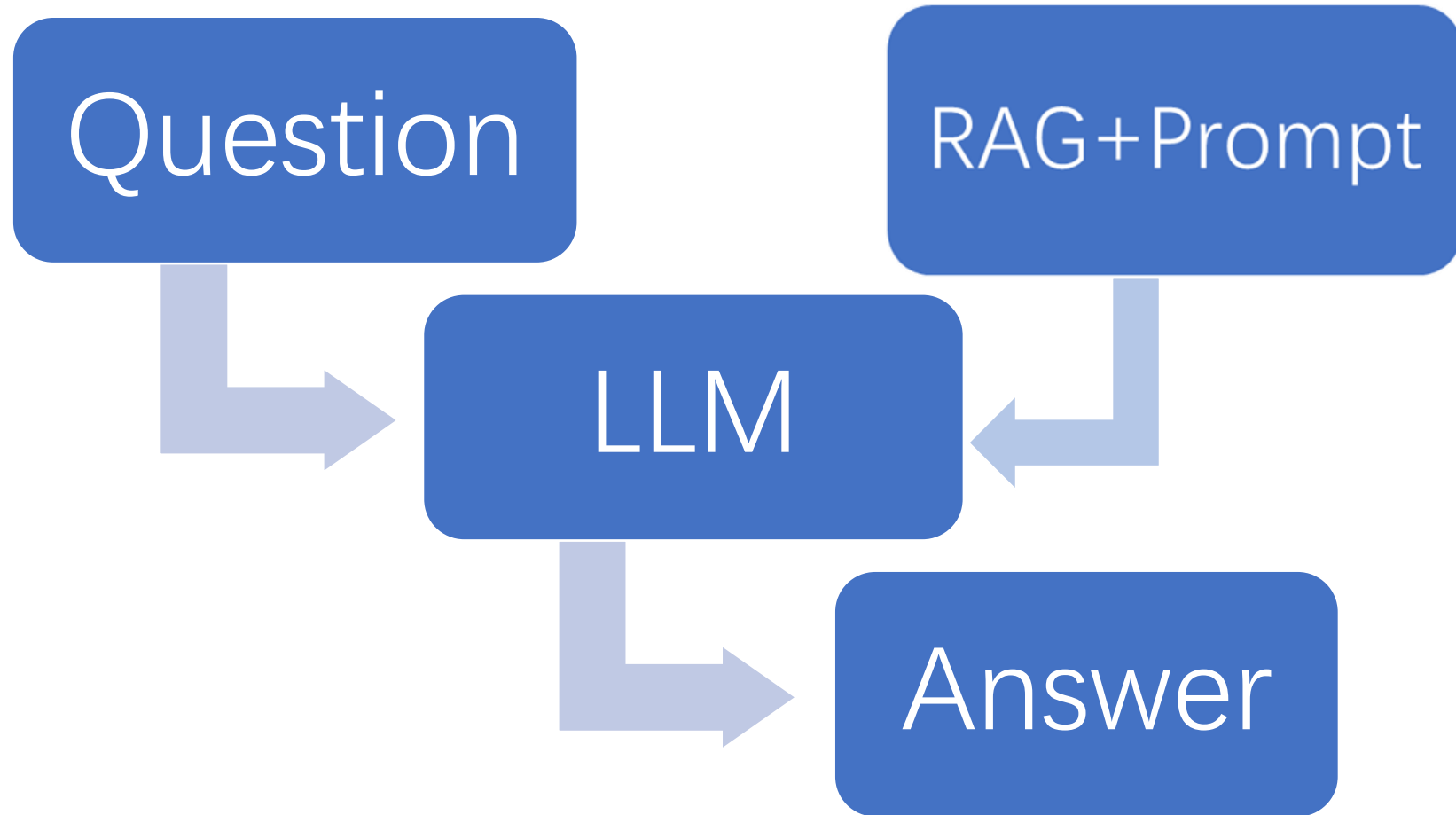
-> RAG 直接检索输入问题

-> 取消检验直接生成

- 思维链过长，自我反思过多，生成答案过慢

-> 使用 prompt engineering 引导

# RAG+Prompt



# Prompt

You are a physics expert. Please solve the following problem. Please use `\\boxed{\\{\\}}` around your final result. **Be confident and don't rethink too much.**

For simple problem, you can **solve it directly**.

If the problem is complex, you can break it down into subproblems and solve them one by one. Some problems can be broken down into four steps: **"rigorously analyze the geometric and physical context", "list equations and transform them into mathematical problems", "solve the equations", and "get conclusions from the results"**.

If the calculation is complex, you can keep the formula and use wolfram language, enclose the calculation expressions within ``<wolfram>`` and ``</wolfram>`` tags, for example: ``<wolfram>2 + 2</wolfram>``. The wolfram engine will compute the result and replace it in the answer.



# RAG

- 采用 snowflake-arctic-embed-xs 模型
- 使用 faiss 相似性搜索和聚类库，用 IndexIVFFlat 算法来加快检索速度
- 语料来源
  - stackexchange 中与物理相关的内容
  - 老师提供的
    - > 约 22 万条数据

# RAG

```
"1": {
  "Title": "What is spin as it relates to subatomic particles?",
  "Q": "<p>I often hear about subatomic particles having a property called \"spin\" but also that it doesn't actually relate to spinning about an axis like you would think. Which particles have spin? What does spin mean if not an actual spinning motion?</p>\n",
  "A": "<p>Spin is a technical term specifically referring to <em>intrinsic angular momentum</em> of particles. It means a very specific thing in quantum/particle physics. (Physicists often borrow loosely related everyday words and give them a very precise physical/mathematical definition.)</p>\n\n<p>Since truly fundamental particles (e.g. electrons) are point entities, i.e. have no true size in space, it does not make sense to consider them 'spinning' in the common sense, yet they still possess their own angular momenta. Note however, that like many quantum states (fundamental variables of systems in quantum mechanics,) spin is <em>quantised</em>; i.e. it can only take one of a set of discrete values. Specifically, the allowed values of the <a href=\"http://en.wikipedia.org/wiki/Spin_%28physics%29#Spin_quantum_number\" rel=\"noreferrer\">spin quantum number</a> $s$ are non-negative multiples of 1/2. The actual spin momentum (denoted $S$) is a multiple of Planck's constant, and is given by $S = \hbar \sqrt{s(s+1)}$.</p>\n\n<p>When it comes to composite particles (e.g. nuclei, atoms), spin is actually fairly easy to deal with. Like normal (orbital) angular momentum, it adds up linearly. Hence a proton, made of three constituent quarks, has overall spin 1/2.</p>\n\n<p>If you're curious as to how this (initially rather strange) concept of spin was discovered, I suggest reading about the <a href=\"http://en.wikipedia.org/wiki/Stern-Gerlach_Experiment\" rel=\"noreferrer\">Stern-Gerlach experiment</a> of the 1920s. It was later put into the theoretical framework of quantum mechanics by Schrodinger and Pauli.</p>\n",
  "Score": "52"
},
"2": {
  "Title": "What is your simplest explanation of string theory?",
  "Q": "<p>How would you explain string theory to non-physicists such as myself? I'm especially interested in how plausible is it and what is needed to successfully prove it?</p>\n",
  "A": "<p>I've noticed that none of these answers actually answer the question.</p>\n\n<p>The simplest explanation of string theory I can think of:</p>\n\n<p><strong>Particles we currently consider \"point particles\" (electrons, quarks, photons, etc.) are actually tiny pieces of string with each a characteristic vibration. They interact in a sort of harmony that results in/manifests as the physical laws we observe.</strong></p>\n\n<p>If anyone with more knowledge in the field can correct me, I ask for improvements. This is just how I personally explain it to people who ask, and I'd hate to give out false information.</p>\n",
  "Score": "14"
},
```

# RAG

```
9  ## Problem 2.1
10
11  A point charge  $q$  is brought to a position a distance  $d$  away from an infinite plane conductor held at zero potential. Using the method of images, find:
12  (a) the surface-charge density induced on the plane, and plot it;
13  (b) the force between the plane and the charge by using Coulomb's law for the force between the charge and its image;
14  (c) the total force acting on the plane by integrating  $\sigma^2 / 2 \epsilon_0$  over the whole plane;
15  (d) the work necessary to remove the charge  $q$  from its position to infinity;
16  (e) the potential energy between the charge  $q$  and its image (compare the answer to part d and discuss).
17  (f) Find the answer to part d in electron volts for an electron originally one angstrom from the surface.
18  (a) We'll take  $d$  to be in the  $z$  direction, so the charge  $q$  is at  $(x, y, z) = (0, 0, d)$ . The image charge is  $-q$  at  $(0, 0, -d)$ . The potential at a point  $\mathbf{r}$  is
19
20  $$
21  \Phi(\mathbf{r}) = \frac{q}{4 \pi \epsilon_0} \left[ \frac{1}{|\mathbf{r} - d \mathbf{k}|} - \frac{1}{|\mathbf{r} + d \mathbf{k}|} \right]
22  $$
23
24  The surface charge induced on the plane is found by differentiating this:
25
26  $$
27  \begin{aligned}
28  \sigma &= -\epsilon_0 \frac{\partial \Phi}{\partial z} \Big|_{z=0} \\
29  &= -\epsilon_0 \frac{q}{4 \pi} \left[ \frac{-(-d)}{r^3} - \frac{(d)}{r^3} \right] \Big|_{z=0} \\
30  &= -\frac{q d}{2 \pi} \frac{1}{(x^2 + y^2 + d^2)^{3/2}} \tag{1}
31  \end{aligned}
32  $$
33
34  We can check this by integrating this over the entire  $x y$  plane and verifying that the total charge is just the value  $-q$  of the image charge:
35
36  $$
37  \begin{aligned}
38  \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \sigma(x, y) dx dy &= -\frac{q d}{2 \pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{1}{(r^2 + d^2)^{3/2}} dx dy \\
39  &= -q d \int_{-\infty}^{\infty} \frac{1}{r} \frac{1}{(r^2 + d^2)^{3/2}} dx \\
40  &= -\frac{q d}{2} \int_{-\infty}^{\infty} \frac{1}{u^2} u^{-3/2} du \\
41  &= -\frac{q d}{2} \left[ -\frac{1}{u^{1/2}} \right]_{-\infty}^{\infty} \\
42  &= -q \quad \text{quad} \quad \frac{1}{\sqrt{u}} \\
43  \end{aligned}
44  $$
45
46  (b) The point of this problem is that, for points above the  $z$  axis, it doesn't matter whether there is a charge  $-q$  at  $(0, 0, -d)$  or an infinite grounded sheet at  $z=0$ . Physics above the  $z$  axis is exactly the same whether we have the charge or the sheet. In particular, the force on the original charge is the same whether we have the charge or the sheet. That means that, if we assume the sheet is present instead of the charge, it will feel a reaction force equal to what the image charge would feel if it were present instead of the sheet. The force on the image charge would be just  $F = q^2 / (4 \pi \epsilon_0 d^2)$ , so this must be what the sheet feels.
```

# 测试比对

1. 大模型的知识储备非常完善，几乎很少用到rag来辅助；并且每次用到rag的时候，会尝试与自己现有的知识库进行匹配，直到两者“match”，它才会放心使用

Okay, let me try to figure out how to solve this problem step by step. So, we have a point charge  $q$  near an infinite grounded conducting plane. The charge is released from rest at a distance  $d$  and moves towards the plane. I need to find the time  $T$  it takes to reach the plane.

First, I remember that when dealing with charges near conductors, image charges come into play. From document 2, I see that the image charge method was used, where the conducting plane has an image charge  $-q$  located symmetrically behind the plane. This setup creates an effective force on the original charge.

In document 1, it was mentioned that the potential in region  $V'$  (the region beyond the conducting plane) is zero because the induced surface charge density  $\sigma$  is zero. But wait, actually, in document 2, they calculated the surface charge density as  $\sigma = -(q d)/(2\pi(r^2 + d^2)^{3/2})$  which isn't zero everywhere except at infinity. Hmm, maybe I misread document 1. Anyway, moving on.

Since the charge is moving under the influence of the electrostatic force, I should probably consider the forces involved. The force on the charge  $q$  is due to its interaction with the conducting plane. From document 2, part (b), the force on the charge is  $F = q^2/(16 \pi \epsilon_0 d^2)$ . Wait, no, that was the force on the image charge, but since the conducting plane exerts a force equal and opposite, the actual force on  $q$  is  $F = q^2/(16 \pi \epsilon_0 d^2)$ .

But hold on, in reality, the force on the charge  $q$  due to the conducting plane can also be thought of as the attraction between  $q$  and its image charge  $-q$ . The distance between them is  $2d$ , so the force should be  $F = (1/(4 \pi \epsilon_0)) * (q * (-q))/( (2d)^2 ) = q^2/(16 \pi \epsilon_0 d^2)$ . Yeah, that matches.

Now, knowing the force, I can relate it to acceleration. Since  $F = ma$ , then  $a = F/m = q^2/(16 \pi \epsilon_0 m d^2)$ . But wait, the force isn't constant because as the charge moves closer to the plane,  $d$  decreases, so the force increases. Therefore, the acceleration isn't constant, making this a variable acceleration problem.

2. 但是我们发现大模型通常会过度自信，它似乎坚信自己能够不依赖外界工具的帮助独立完成作业。例如，我们要求模型在遇到数值计算的时候，借助wolfram的帮助，但是模型从没有生成过任何mma的公式；并且，似乎rag的参考作用有限，我们发现有些题目在调用rag后可以很轻松的解决，但是模型却无法得到正确答案

For northern hemisphere:  $\int_0^{\pi/2} \sin^3\theta \, d\theta = (3/8)\pi$

For southern hemisphere:  $\int_{\pi/2}^{\pi} \sin^3\theta \, d\theta = \text{also } (3/8)\pi$ , because  $\sin(\pi - \theta) = \sin\theta$ .

Similarly, because  $r^2 = x^2 + y^2 + z^2$ , so  $r^2 - 6x^2 = y^2 + z^2 - 5x^2$ . **Wait**, no, hold on. Actually,  $r^2$  is  $x^2 + y^2 + z^2$ , so if I factor out  $(r^2)^{-5/2}$ , it becomes:

Another approach is to consider the magnetic pressure. The internal magnetic pressure in a material can lead to forces. The magnetic pressure is given by  $P = (\mu_0 / 3) B^2$ . But in this case, the sphere isn't a material; it's a collection of charges. Hmm.

**Wait**, another thought: when the sphere rotates, it creates a magnetic field similar to that of a dipole. The field outside the sphere is that of a dipole, while inside, it's different. But since we're dealing with the force between hemispheres, which are part of the same sphere, maybe we need to look at the internal field.

The internal magnetic field of a dipole is given by  $B_{\text{internal}} = (\mu_0 \mu) / (4\pi) * (3 \cos\theta \hat{y} - \hat{u}) / r^3$ , but I might be mixing things up. Alternatively, the internal field of a uniformly magnetized sphere is uniform and given by  $B = (2\mu_0 M) / (3)$  inside, where  $M$  is the magnetization. But in our case, it's an electric dipole due to rotation, so maybe the analogy holds.

to the integral, but the total integral is  $4/3 \approx 1.333$ , whereas  $(3/8) * 2\pi \approx 1.885$ , which is larger than  $4/3$ . **Wait**, something's wrong here.

**Wait**, no,  $\int_0^{\pi} \sin^3\theta \, d\theta = 4/3 \approx 1.333$ . If I split it into two equal parts, each from 0 to  $\pi/2$  and  $\pi/2$  to  $\pi$ , each integral would be  $2/3 \approx 0.666$ . Because  $\sin^3\theta$  is symmetric around  $\theta = \pi/2$ , so each half contributes half the total integral.

at **wait**, the problem doesn't specify whether the inner cavity is filled with anything. It just says it's a hollow dielectric sphere. So perhaps the inner cavity is a vacuum, and the rest is the dielectric. Or maybe the entire sphere except the cavity is dielectric. Hmm, I'll proceed assuming that the cavity is empty, and the dielectric is only in the shell between inner and outer radii.

尝试检查数学计算的正确性

尝试从另一个视角/思路来解决问题

尝试整理自己的思路，思考下一步怎么做

Wait, let's do it properly:

Inside  $\sqrt{\epsilon_0 m}$ :  $\epsilon_0$  has units  $C^2 s^4 kg^{-1} m^{-3}$ ,  $m$  is  $kg$ . So  $\epsilon_0 m$  has units  $C^2 s^4 kg^{-1} m^{-3} * kg = C^2 s^4 m^{-3}$ .

Integrating from 0 to  $\pi$ , we get  $\mu = \sigma \pi R^4 \omega \int_0^\pi \sin^3 \theta d\theta = \sigma \pi R^4 \omega * (4/3)$ .

Wait, actually,  $1/d - 1/x = (x - d)/(dx)$ . So  $\sqrt{1/d - 1/x} = \sqrt{(x - d)/(dx)}$ . But since  $x$  goes from 0 to  $d$ ,  $(x - d)$  is negative, so taking square root of a negative number isn't real. Oops! That suggests I made a mistake in the sign.

Wait, no, because when  $x$  approaches  $d$ ,  $(x - d)$  approaches 0, but for  $x < d$ ,  $(x - d)$  is negative. So  $\sqrt{\text{negative}}$  isn't real. That indicates an error in the setup. any constants. So, perhaps the issue is elsewhere.

$T$  has units  $(C s^2 m^{-3/2}) / C = s^2 m^{-3/2}$ .

Wait, that doesn't match  $[s]$ . Something's wrong.

4.太多的自我思考和自我反思会极大延长模型生成问题的时间，降低生成效率——我们在查看几组的answer后发现，很多组在一些问题上生成了一半就被max\_token强行终止，我们怀疑最后一道题甚至是因为到规定时间而被强行终止；

所以需要利用rag和mma等外部工具来代替或者加速模型的一部分思考，实现效率和生成质量/严谨性的平衡



field inside the conductor is negligible. However, the field outside the  
previously established. **Wait**, but for the purpose of self-inductance, we need to find the flux through the  
area enclosed by the outer conductor due to its own current.  
But the outer conductor is a hollow cylinder,  
so the flux through its area is the integral of its magnetic field over the annular region from  $a$  to  $b$ .  
**Wait**, no. The outer conductor's current produces a magnetic field outside itself, but the flux through  
the area bounded by the outer conductor is zero because the field lines outside the conductor do not pass  
through the annular region between  $a$  and  $b$ . **Wait**, that might not make sense. Let me think again.  
When calculating the self-inductance of the outer conductor, you imagine the current going through it and compute  
the flux through the area enclosed by the conductor. But since the outer conductor is a thin shell, the flux  
through its own area (the annulus between  $a$  and  $b$ ) is zero because the magnetic field inside the conductor is  
zero (as it's a surface current) and outside is negligible for the purposes of flux through the annulus.  
**Wait**, no, actually, the magnetic field outside the outer conductor is zero, but the flux through the  
annulus (between  $a$  and  $b$ ) is due to the outer conductor's current. **Wait**, no, the outer conductor's current  
produces a magnetic field outside itself, but the annulus is inside the outer conductor. So, the magnetic  
field inside the outer conductor (which is the region between  $a$  and  $b$ ) is due to the inner conductor's  
current only, because the outer conductor's current is on its surface at radius  $b$ .  
Therefore, the flux through the annulus (from  $a$  to  $b$ ) due to the outer conductor's current is zero because the outer conductor's  
current doesn't create a field in that region—it's on the boundary.  
Therefore, the self-inductance of the  
outer conductor is zero? That can't be right. **Wait**, perhaps I'm misunderstanding. The self-inductance of a  
loop is the flux through the loop when current passes through it. For the outer conductor, which is a loop at  
radius  $b$ , the flux through it is due to its own current. But since the outer conductor's current is on its  
surface, the magnetic field inside the conductor (at radius  $b$ ) is zero. Therefore, the flux through the loop  
is zero, implying that the self-inductance of the outer conductor is zero.  
That seems odd, but  
mathematically, if the current is on the surface, the field inside the conductor is zero, hence no flux  
through the loop. So,  $L_2 = 0$ .  
Is that possible? It feels counterintuitive because the outer conductor  
still has current, but perhaps because the current is on the surface, the flux through the loop is zero.  
Alternatively, maybe I should model the outer conductor as a solenoid? No, it's a single turn, not  
multiple turns. **Wait**, perhaps I made a mistake earlier. Let me re-examine.  
For a single loop of radius  $R$  carrying current  $I$ , the magnetic field at the center is  $B = \mu_0 I / (2R)$ . The flux through the loop is  $\Phi = B \cdot$

我们反思了考试题目的错因：

P3：在做复杂积分的时候出错，参数d被意外"canceled out"

（可能可以通过调用mathematica解决）

P4：题目建模对于模型过于复杂，导致模型在前期理解问题花费大量时间，最终因为到达max\_token而停止作答

P7：并未充分利用题目条件，导致方法上出现错误（可能可以通过调用RAG解决）

P8：也是未能充分利用题目条件，猜测可能是题目对于模型过于复杂

# 于是，我们采用了“满血版”的模型进行测试

- condition：我们放宽了max\_token和限制作答时间的限制
- result：
- 之前做对的题目依然能够做对
- P3：可以做对
- P4：相比考试答案更近一步，仅仅和标准答案相差一个常数，出错在积分运算
- P7：模型已经清晰列出了需要求解的方程，但是它觉得“don't have enough time to figure it out”，遂放弃，但是相比之前仍然进步很多
- P8：依然无法战胜，推测是此类问题见得不多，模型对此的泛化性能不能帮助他解决问题

# 不足与展望

- R1 有时会根据训练时已有的信息，直接“背”出结果，而不是经过严谨的推理得到结果
  - 思维链仍然过长，即使经过多次反思也无法得到正确结果
  - R1 倾向于思考的时候算完整个题目，且计算容易算错
- > 微调模型，使其在处理问题时先理顺思路，再列式求解；在需要计算结果判断思路是否正确或可以继续按该思路思考时主动停止，尝试调用外部计算软件来得到结果，然后再进行下一步思考与计算

# 不足与展望

- 电动力学问题的核心挑战常在于边界条件（如导体表面、介质界面），模型可能无法准确关联物理场景与数学边界条件设置
  - RAG检索的文献中若缺少类似案例，模型难以通过类比推理生成正确解
- 电动力学依赖高度形式化的数学语言（如矢量微积分、张量、边界条件），模型可能并不擅长使用这类语言
  - 模型对数学符号的理解本质上是统计模式匹配，而非严格数学推理
- 难以像人类一样通过物理直觉简化问题（如对称性分析）
  - 模型依赖文本训练数据，本质还是预测下一个token，缺乏对物理本质的认知

谢谢大家！