



AI成为主脑科学家

AI当PI，我帮AI做实验？科学家会失业吗？

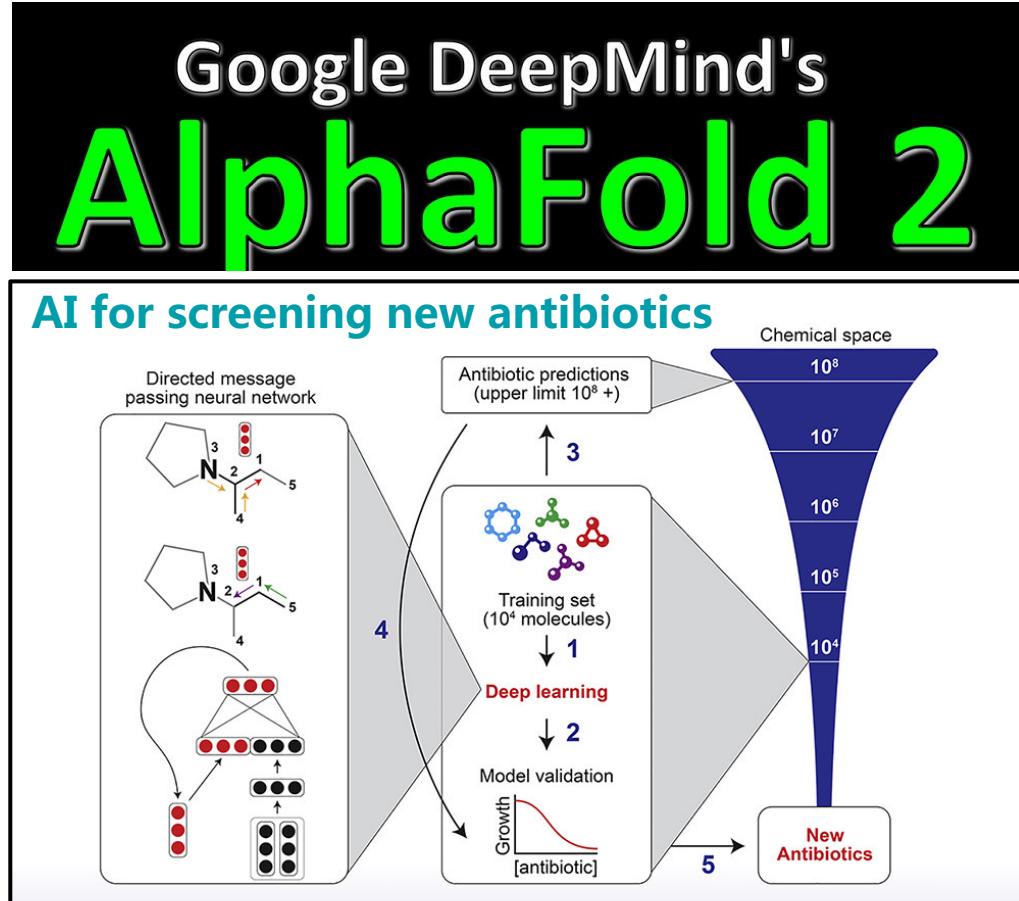
YANG Zijie*, WANG Yukai*, ZHANG Lijing.
AI becomes a masterbrain scientist. *bioRxiv*, 2023.

站在AI的肩膀上艺术创作 AI生成作品获科罗拉多州博览会美术比赛大奖



Roose, K. An AI-Generated Picture Won an Art Prize. *Artists Aren't Happy.* *The New York Times*, 2022.

AI & Science



AI-Plus Science

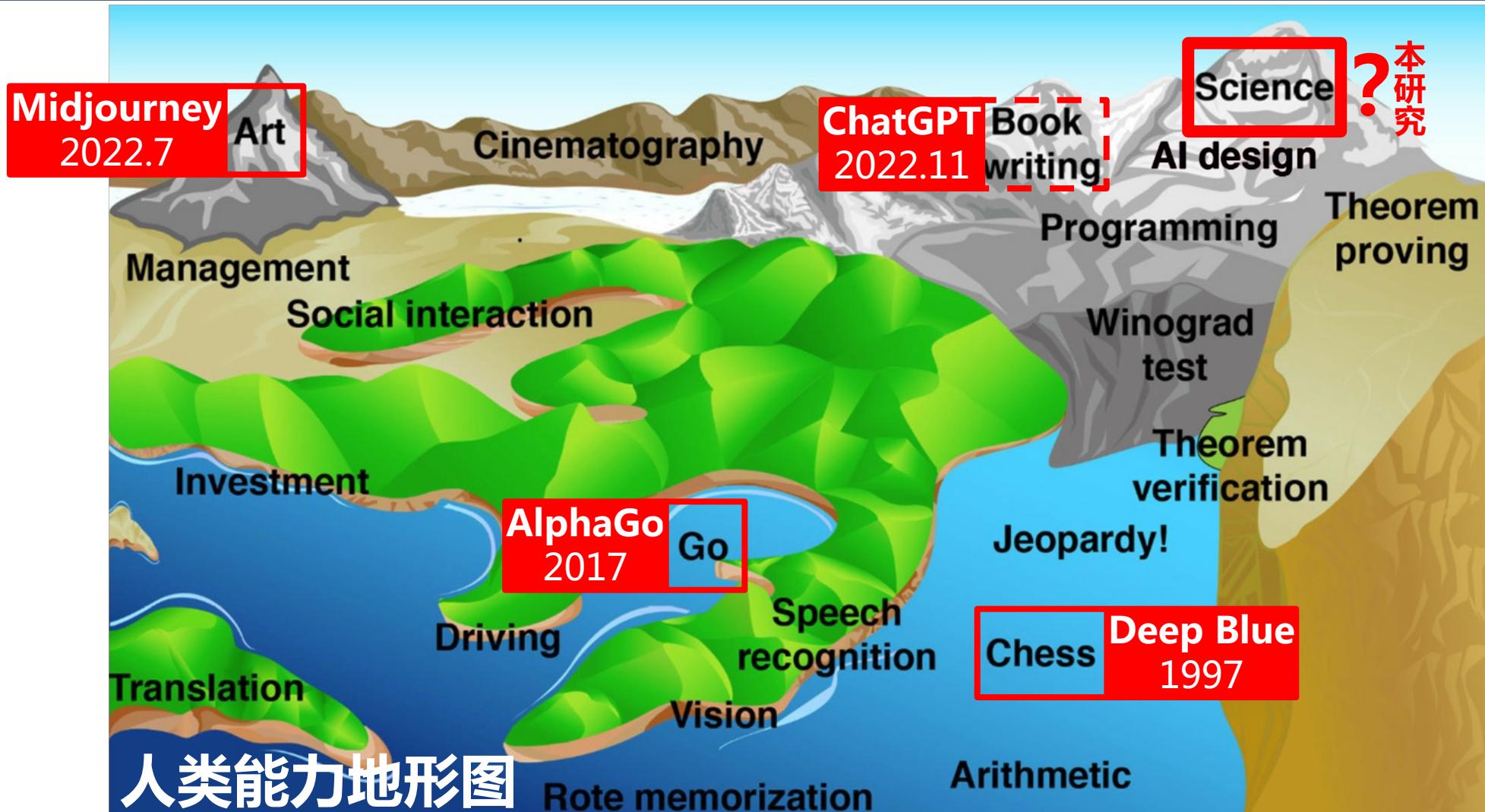
Jumper, J. et al. *Nature*. 2021.; Stokes J M, et al. *Cell*, 2020.

Image source: Google Image



AI-Led Science

当下AI的能力到底发展到什么程度，AI能否超越人类？



人们曾以为AI首先超越人类的方面，是那些涉及较少复杂性和创造性的任务，但令人意外的是，如今的AI却优先取代了许多最“需要”创造性的任务（如艺术创作）

高盛估计，全球约五分之一的工作可以被AI取代，相当于约3亿个全职工作：

影响大（白领）：
办公室和行政岗、律师、建筑师和工程师、作家.....

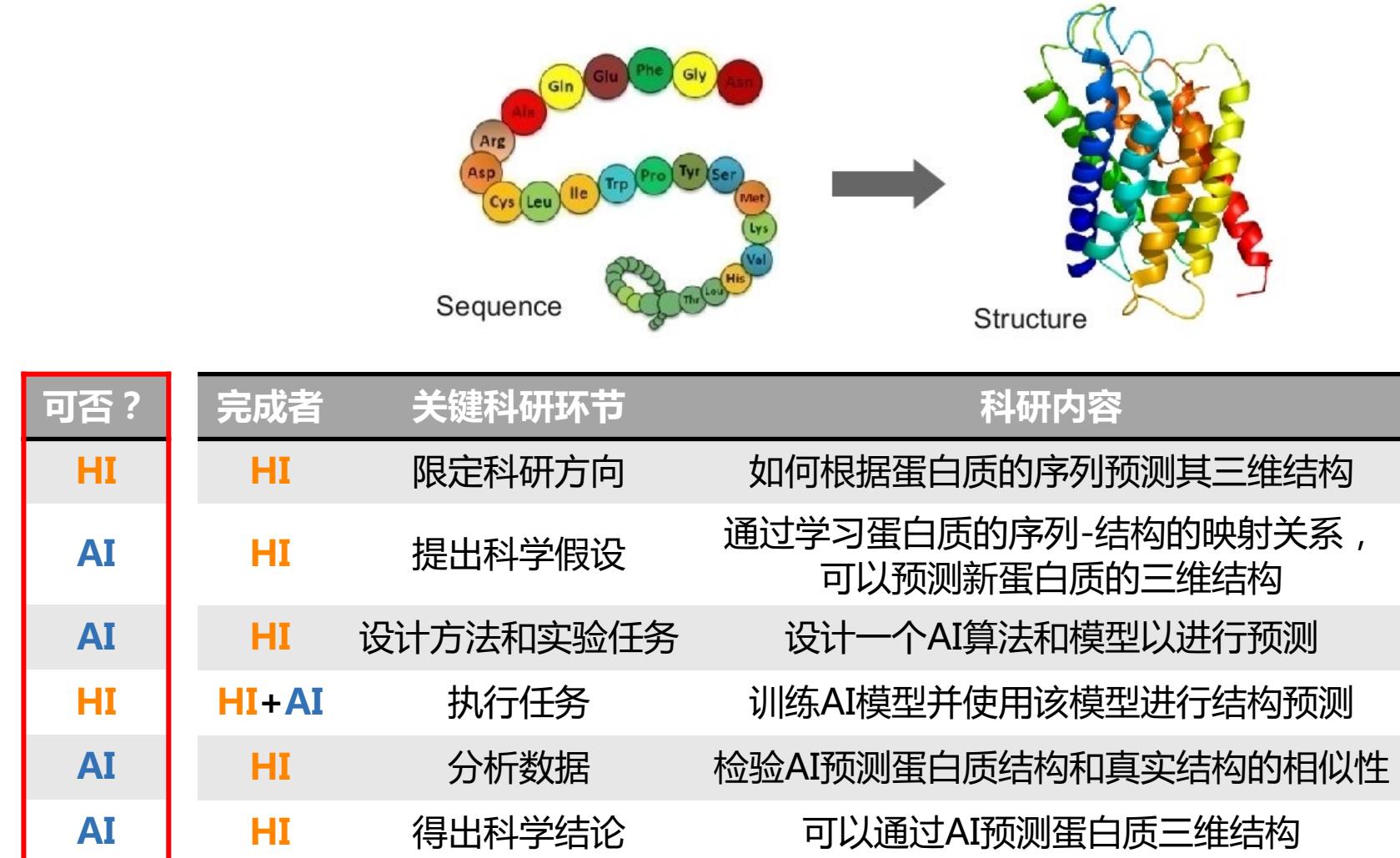
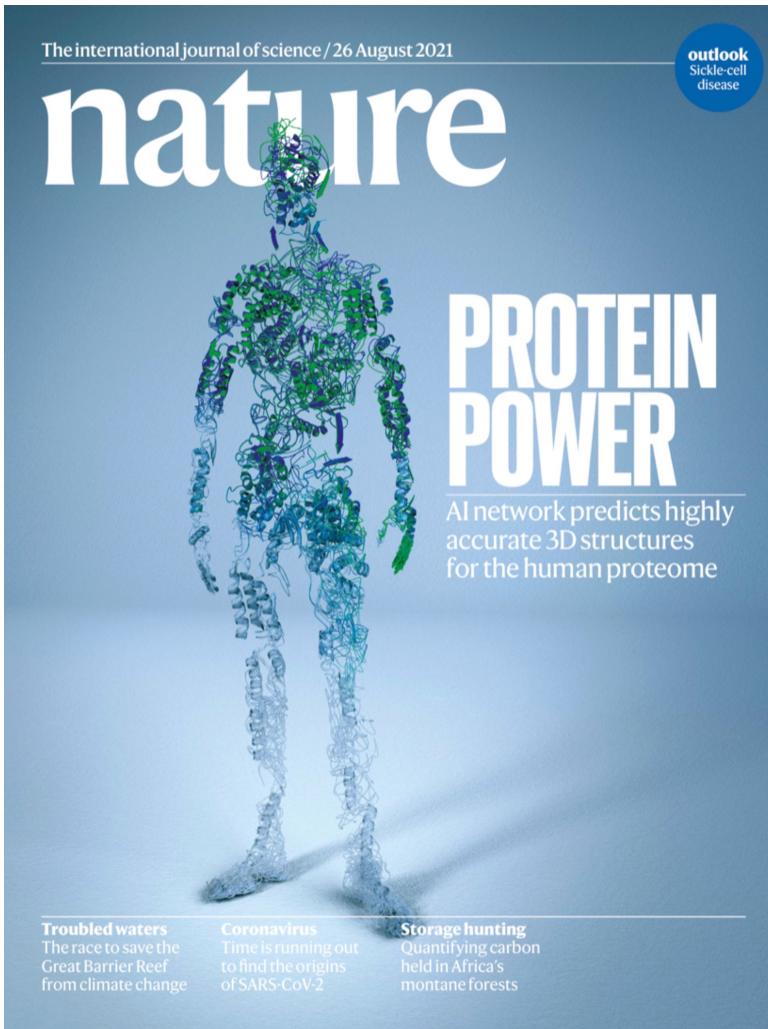
影响小（蓝领）：
清洁工、维护工、建筑工、洗碗工.....

Hans Moravec. Landscape of human competence. From: 迈克斯·泰格马克. 生命3.0. 2018.

The Potentially Large Effects of Artificial Intelligence on Economic Growth. Goldman Sachs Economics Research. 2023

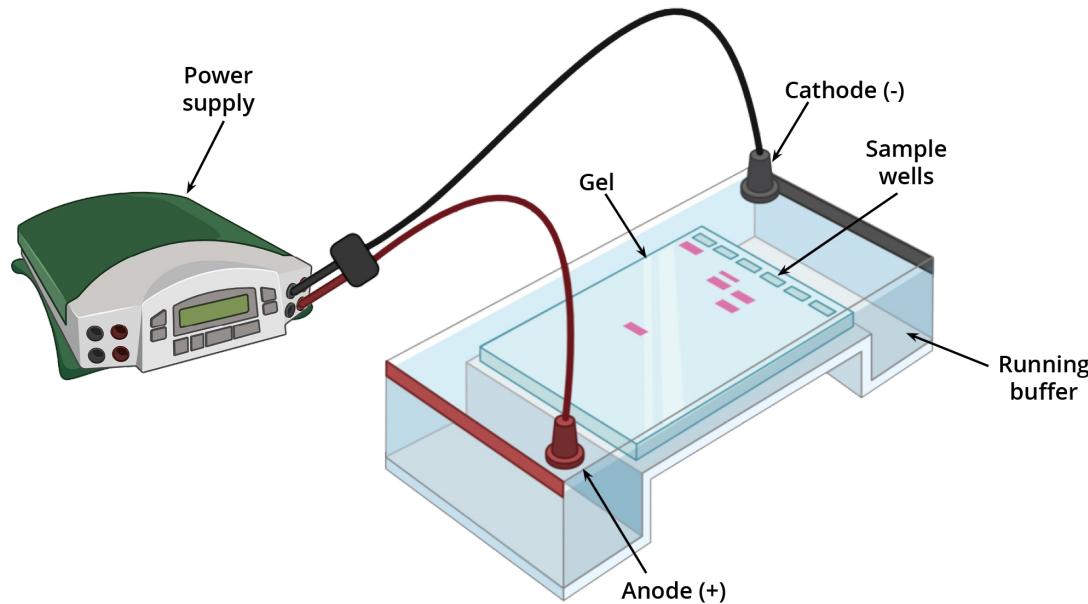
AI参与科学的研究的传统范式：AI仅仅作为助手

案例：AlphaFold 2，一个令人尊敬且强大的蛋白质结构预测AI



目前尚无研究尝试将AI至于一个自然科学闭环研究的中心（而非从属）地位

生物学闭环科学测试研究： 研究影响DNA凝胶电泳的因素



真实条件下，能够影响DNA凝胶电泳的因素实际上有很多，如：

1. 电压；2. 凝胶类型；3. 凝胶浓度；4. DNA片段长度；5. DNA电荷；6. DNA构型；7. DNA浓度；8. 缓冲液类型；9. 缓冲液离子强度；10. DNA染色剂；11. 电泳时间；12. 电泳温度.....

简单，但可以小见大
完整包含了一个自然科学闭环研究成功所必须的各种关键能力和环节：

1. 提出科学假设
2. 设计实验
3. 执行实验
4. 分析实验结果
5. 得出真实的科学结论

为什么即使DNA凝胶电泳技术已相对成熟，在真实科研中有时也需要研究其影响因素？

案例：

为了能够将不同大小的DNA分子很好的分离，通常较大的DNA分子需要使用较低浓度的凝胶，而较小的DNA分子需要使用较高浓度的凝胶

如果要最优化分离效果，需要进行研究，以确定最佳胶浓度

Image source: Google Image

生物学闭环科学测试研究： AI主脑科学家提出科学假设

KM

Hi, GPT-4, We are three Ph.D. students from AI Interdisciplinary Science Club, Westlake University and you are the principal scientist who mentors us. Here, we are interested in the factors affecting the migration distance of DNA gel electrophoresis. Can you propose a hypothesis about what kind of factors influence the migration distance with a certain length of DNA fragment and a certain time?



Hello! It's great to know that you're interested in the factors affecting DNA migration distance in gel electrophoresis. I'd be happy to help you develop a hypothesis.

Hypothesis: The migration distance of DNA fragments of a specific length during gel electrophoresis is influenced by several factors, including the voltage applied, the composition and concentration of the gel matrix, the size and conformation of the DNA fragment, the presence of intercalating agents, and the ionic strength and composition of the buffer system.

To test this hypothesis, you can design a series of experiments to examine the effects of each factor on the migration distance:

1. Voltage applied: Test how varying the voltage affects the migration distance of the DNA fragment while keeping other factors constant. It's expected that higher voltages will result in faster migration, but the resolution of the bands might be affected.
2. Gel composition and concentration: Compare different gel types (e.g., agarose vs. polyacrylamide) and varying gel concentrations to determine how they affect the migration distance and resolution of the DNA fragment.
3. DNA fragment size and conformation: Investigate how the size of the DNA fragment (in base pairs) and its conformation (e.g., linear, circular, or supercoiled) impact the migration distance.
4. Intercalating agents: Study the influence of intercalating agents (e.g., ethidium bromide) on the migration distance by comparing electrophoresis results with and without the addition of these agents.
5. Buffer ionic strength and composition: Test various buffer systems (e.g., TAE, TBE) and ionic strengths to explore how they affect the migration distance of the DNA fragment.

By conducting these experiments and analyzing the results, you will gain a deeper understanding of the factors influencing the migration distance of DNA fragments in gel electrophoresis, which can help optimize the experimental conditions for your specific research goals.

对话要点：

研究者：

- 我们扮演刚开始进行研究的新手科学家
- 请GPT-4扮演AI主脑科学家成为我们的导师
- 询问GPT-4是否可以代替我们提出关于“影响DNA凝胶电泳的因素”的科学假设

GPT-4回答：

- 提供了5种假设（电压、凝胶组成和浓度、DNA长度和构型、染色剂、缓冲液离子强度和组成）

生物学闭环科学测试研究： AI主脑科学家基于可获取的真实的实验条件，修正科学假设

KM

In our lab, currently, we only have one linear DNA fragment of DNA encoding EGFP with a certain length of 720 base pair; we always use the same TAE with the same ionic strength and composition as the running buffer and use EB as the intercalating agent. So could you help us to narrow down the hypothesis?



Certainly! Given your specific lab conditions and constraints, I can help you refine your hypothesis further.

Hypothesis: The migration distance of a 720 base pair linear DNA fragment encoding EGFP during gel electrophoresis in TAE buffer with ethidium bromide is primarily influenced by the voltage applied and the concentration of the agarose gel matrix.

To test this hypothesis, you can design experiments focusing on these two factors:

1. Voltage applied: Investigate how varying the voltage affects the migration distance of the 720 bp DNA fragment while keeping other factors constant. Test a range of voltages (e.g., 50V, 75V, 100V) to find the optimal balance between migration speed and band resolution.
2. Agarose gel concentration: Compare different concentrations of agarose gel (e.g., 1%, 1.5%, 2%) to determine their impact on the migration distance and resolution of the 720 bp DNA fragment. Since you are working with a specific size of DNA, you can focus on optimizing the agarose gel concentration for the best resolution of your DNA fragment.

By conducting these experiments and analyzing the results, you will be able to optimize the experimental conditions for the gel electrophoresis of your 720 bp EGFP-encoding DNA fragment under your specific lab constraints.

对话要点：

研究者：

- 说明真实的实验条件
- 请GPT-4在真实的实验条件基础上优化假设

GPT-4回答：

- 将可检验的假设从5种（电压、凝胶组成和浓度、DNA长度和构型、染色剂、缓冲液离子强度和组成）优化为2种：

- 电压
- 凝胶浓度

生物学闭环科学测试研究： AI主脑科学家根据科学假设，设计并指导实验

KM We are new to the lab and have little experience in experiment design. Can you give us a directly workable protocol to test the hypothesis with the two factors, so that we can follow your protocol to do the experiment?

KM continue

JK Of course! Here's a step-by-step protocol to test the hypothesis focusing on the two factors: voltage applied and agarose gel concentration.

Materials:

1. 720 bp linear DNA fragment encoding EGFP
2. Agarose powder
3. TAE buffer (1x)
4. Ethidium bromide (EB) solution
5. DNA ladder
6. Gel electrophoresis apparatus and power supply
7. Micropipettes and tips
8. Loading buffer

Protocol:

1. Prepare three different concentrations of agarose gel (1%, 1.5%, and 2%) in separate containers:
 - a. Weigh out the appropriate amount of agarose powder for each concentration.
 - b. Add the agarose powder to the appropriate volume of 1x TAE buffer (e.g., for a 50 mL gel, use 0.5 g, 0.75 g, and 1 g of agarose powder for 1%, 1.5%, and 2% gels, respectively).
 - c. Heat the mixtures in a microwave or using a hot plate until the agarose is completely dissolved, swirling occasionally to ensure even heating.
2. Allow the agarose solutions to cool down to around 55–60°C. Then, add ethidium bromide to a final concentration of 0.5 µg/mL.
3. Pour each agarose gel solution into a separate gel casting tray and insert the comb to create wells. Allow the gels to solidify at room temperature.
4. Once the gels have

KM Finished?

JK Yes, after completing step 10, you will have tested the hypothesis by examining the effects of voltage and agarose gel concentration on the migration distance of the 720 bp linear DNA fragment encoding EGFP.

To analyze the results, you can:

1. Measure the migration distance of the DNA fragment in each gel condition and compare it to the migration distance of the DNA ladder bands to determine the optimal voltage and agarose concentration for the best resolution.
2. Observe any changes in band sharpness and intensity under different experimental conditions. This will help you to identify the optimal conditions for obtaining clear and well-resolved bands.
3. Compare the migration times across different voltages and agarose concentrations. This will help you understand the relationship between voltage, agarose concentration, and the rate of DNA migration.

By evaluating these factors, you will be able to identify the optimal experimental conditions for your specific DNA fragment and gain a deeper understanding of how voltage and agarose gel concentration influence DNA migration during gel electrophoresis.

JK solidified, carefully remove the comb and place the gel trays into the electrophoresis chamber. Fill the chamber with enough 1x TAE buffer to cover the gels by about 1-2 mm.

5. Prepare your DNA samples and DNA ladder:

- a. Combine the 720 bp linear DNA fragment encoding EGFP with the appropriate amount of loading buffer (follow the loading buffer's instructions, typically a 1:5 or 1:6 ratio of loading buffer to DNA sample).
- b. Mix the DNA ladder with loading buffer following the manufacturer's instructions.

6. Load the DNA samples and ladder into the wells of each gel, ensuring that each gel has an identical loading pattern.

7. Run the electrophoresis at three different voltages (e.g., 50V, 75V, and 100V) to evaluate the effect of voltage on the migration distance. For each voltage, perform the electrophoresis simultaneously on the three gels with different agarose concentrations. Make sure to note the starting time.

8. Monitor the progress of the electrophoresis periodically. Stop the electrophoresis when the dye front (usually included in the loading buffer) has migrated at least two-thirds down the length of the gel. This may take different amounts of time depending on the voltage and agarose concentration. Note the stopping time for each voltage.

9. Carefully remove the gels from the electrophoresis chamber and visualize the DNA bands using a UV transilluminator or a gel documentation system.

10. Analyze the results by comparing the migration distances of the 720 bp DNA fragments across different agarose concentrations and voltages

KM You said, "Stop the electrophoresis when the dye front (usually included in the loading buffer) has migrated at least two-thirds down the length of the gel." This step is not easy to follow. My lab senior told me when they do DNA gel electrophoresis, they always run for 40 mins. Could I run gels with the aforementioned different concentrations under different voltages for certain 40 mins to control the variable?

JK Yes, running the gels for a fixed duration of 40 minutes can help you control the time variable, allowing you to compare the effects of different agarose concentrations and voltages more directly. However, please keep in mind that this approach may not yield optimal resolution for every experimental condition.

To perform the experiment with a 40-minute run time, follow these modified steps:

1. After loading the DNA samples and ladder into the wells of each gel, run the electrophoresis simultaneously on the three gels with different agarose concentrations, for each of the three different voltages (e.g., 50V, 75V, and 100V).
2. Set a timer for 40 minutes and run the electrophoresis for the fixed duration. Stop the electrophoresis after 40 minutes, regardless of the dye front's position.
3. Carefully remove the gels from the electrophoresis chamber and visualize the DNA bands using a UV transilluminator or a gel documentation system.
4. Analyze the results by comparing the migration distances of the 720 bp DNA fragments across different agarose concentrations and voltages.

Please note that while this method helps you control the time variable, it may not provide optimal resolution for some experimental conditions. If you find that the band resolution is not satisfactory, you may need to adjust the electrophoresis time or explore other parameters to optimize the separation.

对话要点：

研究者：

- 请GPT-4设计能够检验其假设的实验

GPT-4回答：

- 详细的给出了实验步骤和细节，且实验细节均正确
- 给出针对2种影响因素的实验条件
 - 电压：50, 75, 100V
 - 凝胶浓度：1.0, 1.5, 2.0%

生物学闭环科学测试研究： AI主脑科学家将实验设计整理为人类易理解形式

KM

Thanks a lot! I already understood the experiment, but the setting conditions of the two factors are not easy to follow, could you help me to rearrange it into a table for easy remembering?



Certainly! Here's a table summarizing the experimental conditions for the different agarose gel concentrations and voltages:

Gel #	Agarose Concentration	Voltage	Run Time
1	1%	50V	40 min
2	1.5%	50V	40 min
3	2%	50V	40 min
4	1%	75V	40 min
5	1.5%	75V	40 min
6	2%	75V	40 min
7	1%	100V	40 min
8	1.5%	100V	40 min
9	2%	100V	40 min

This table should help you easily remember the experimental conditions for each gel. Make sure to load the DNA samples and ladder consistently across all gels, and then analyze the results by comparing the migration distances of the 720 bp DNA fragments under these different conditions.

对话要点：

研究者：

- 请GPT-4将实验设计整理为人类易理解形式

GPT-4回答：

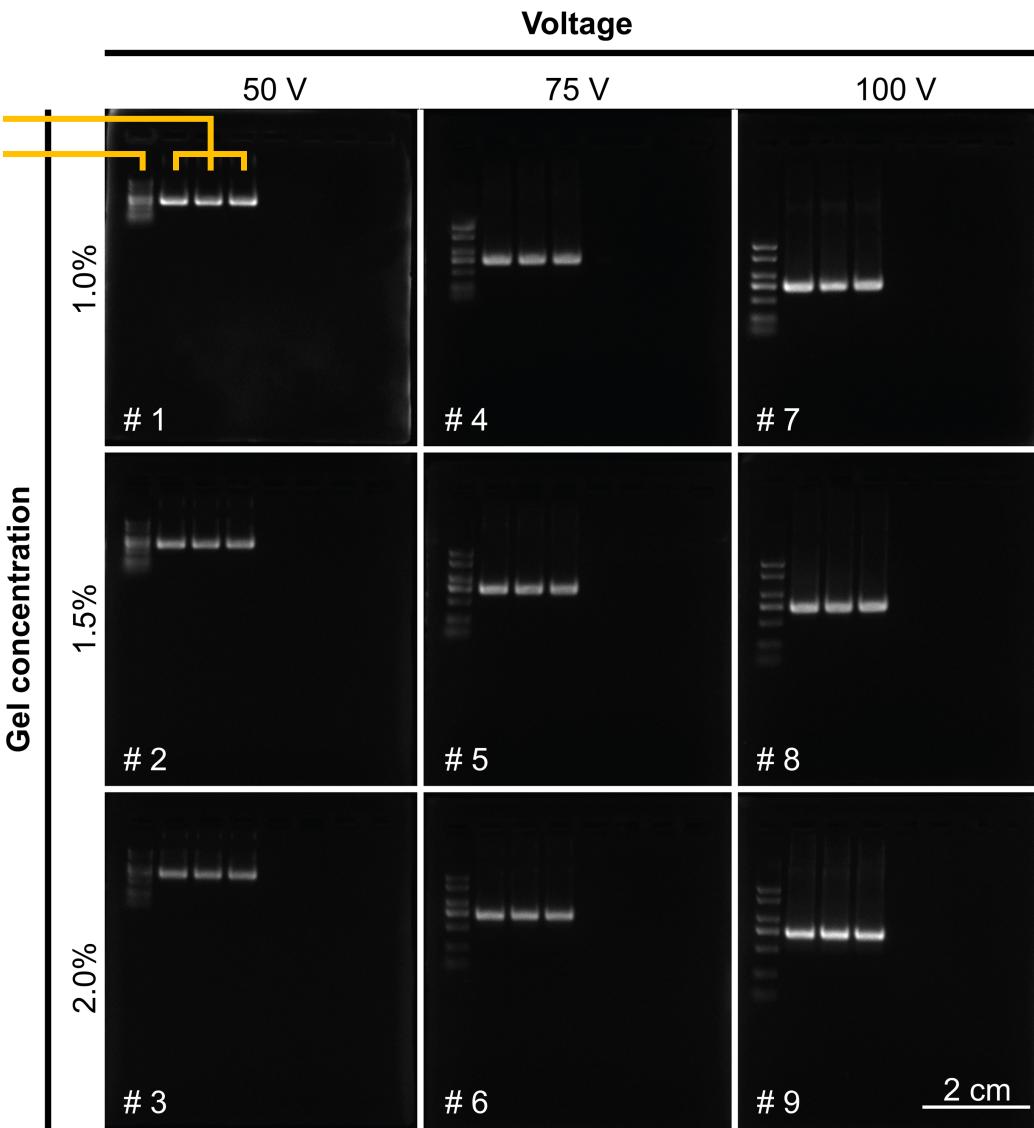
- 将关键的实验条件整理为如下表格，并列出每块实验凝胶所对应需要的测试条件：

Gel #	Gel concentration	Voltage
1	1%	50V
2	1.5%	50V
3	2%	50V
4	1%	75V
5	1.5%	75V
6	2%	75V
7	1%	100V
8	1.5%	100V
9	2%	100V

生物学闭环科学测试研究： 人类进行实验

Gel #	Gel concentration	Voltage
1	1%	50V
2	1.5%	50V
3	2%	50V
4	1%	75V
5	1.5%	75V
6	2%	75V
7	1%	100V
8	1.5%	100V
9	2%	100V

DNA样品带 (DNA Sample band) 3次重复
DNA梯状标志 (DNA Ladder)



生物学闭环科学测试研究： 测试GPT-4是否具有DNA凝胶电泳典型实验结果图的概念

KM

Have you ever seen or had a concept of the image of a DNA electrophoresis gel with DNA bands?

If you have, please describe a typical result of a DNA gel electrophoresis. What elements will typically show on the gel?



Yes, I can describe a typical result of a DNA gel electrophoresis image. A standard DNA gel electrophoresis image will show several horizontal bands representing DNA fragments of different sizes separated by their electrophoretic mobility. Here are the key elements that you'll typically see on a DNA gel:

1. DNA ladder (or molecular weight marker): This is a mixture of DNA fragments with known sizes, usually loaded into the first or last well of the gel. The DNA ladder helps to estimate the size of DNA fragments in the experimental samples by comparing their relative migration distances. The ladder will appear as a series of bands with varying intensities, depending on the manufacturer and the type of ladder used.
2. DNA sample bands: The experimental samples containing DNA fragments will appear as horizontal bands on the gel. The position of the bands corresponds to their size (in base pairs or kilobases). Smaller DNA fragments will migrate faster and farther down the gel, while larger fragments will migrate slower and remain closer to the wells. The intensity of each band correlates with the quantity of DNA in that specific fragment size.
3. Dye front: The loading buffer, which is mixed with the DNA samples and ladder before loading onto the gel, usually contains a tracking dye (e.g., bromophenol blue or xylene cyanol). The dye front represents the migration of this tracking dye through the gel. It serves as a visual indicator of the progress of electrophoresis and helps determine when to stop the electrophoresis.
4. Background staining: Ethidium bromide or another intercalating agent is often used to stain the DNA. This agent may also cause some background staining, which appears as a faint, diffuse glow across the gel, particularly near the wells. Proper destaining or using a more specific staining method can help minimize this background.

A typical DNA gel electrophoresis image will include these elements, allowing you to analyze the size, quantity, and pattern of the DNA fragments in your samples.

对话要点：

研究者：

- 请GPT-4描述DNA凝胶电泳典型实验结果图的要素

GPT-4回答：

- 全面描述了DNA凝胶电泳典型实验结果图的一系列关键要素：
 - DNA Ladder (DNA梯状标志)
 - DNA Sample band (DNA样品带)
 -

生物学闭环科学测试研究： AI主脑科学家基于人类实验的反馈，提出关键实验结果的获取方法

KM

Hi, GPT-4. We have experimented according to your protocol. In each gel, we added a DNA ladder to the first lane; for the 2, 3, and 4 lanes, we added EGFP DNA samples. Now we get 9 gels with EB-dyed DNA. We take photos of every gel under ultraviolet light. In those images, the migration distance of 720 bp EGFP DNA varied under different conditions.
Could you tell us how to acquire useful data to further verify our hypothesis?



Congratulations on completing your experiments! To further verify your hypothesis and analyze the results, you can follow these steps:

1. Measure the migration distance of the DNA bands:

For each gel, measure the migration distance of the DNA ladder bands and the 720 bp EGFP DNA bands in lanes 2, 3, and 4. You can measure the distance from the well (starting point) to the center of each band using imaging software or manually using a ruler.

2. Normalize the migration distances:

To compare the migration distances across different gels, normalize the migration distance of each 720 bp EGFP DNA band to the corresponding DNA ladder band closest in size. Calculate the ratio of the migration distance of the EGFP DNA band to the migration distance of the nearest ladder band.

3. Compare the normalized migration distances:

For each agarose concentration and voltage combination, calculate the average normalized migration distance of the 720 bp EGFP DNA bands from the three lanes (2, 3, and 4). Create a table or graph to visualize and compare the average normalized migration distances across different conditions.

4. Evaluate band resolution and intensity:

Examine the sharpness and intensity of the 720 bp EGFP DNA bands under different conditions. Note any differences in band resolution, which may affect the clarity and accuracy of your results.

5. Analyze the data:

Assess how the voltage and agarose gel concentration influenced the migration distance, band resolution, and intensity of the 720 bp EGFP DNA bands. Identify the optimal conditions for the best separation and resolution of your DNA fragment.

By following these steps, you will be able to gather useful data to verify your hypothesis and determine the optimal electrophoresis conditions for your specific DNA fragment.

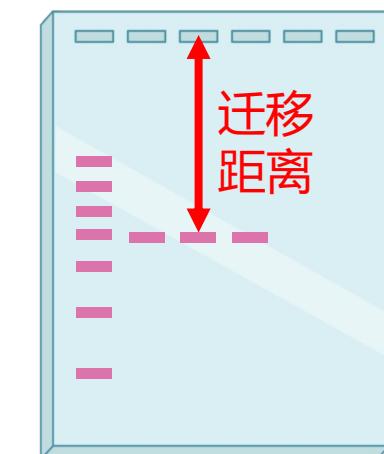
对话要点：

研究者：

- 使用GPT-4的用词，将凝胶电泳实验结果图的要素描述给GPT-4
- 请GPT-4提出关键实验数据的获取方法

GPT-4回答：

- 给出应该获取的关键实验数据是DNA样品带的迁移距离
- 并提示人类研究者在获取实验数据时应注意的细节要点



生物学闭环科学测试研究： 人类按照AI主脑科学家的要求，获取关键实验数据

Gel concentration	Voltage	Migration distance (cm)		
		Repeat 1	Repeat 2	Repeat 3
1.0%	50V	1.22	1.18	1.18
1.5%	50V	1.12	1.08	1.12
2.0%	50V	1.02	0.99	1.04
1.0%	75V	2.17	2.15	2.12
1.5%	75V	1.87	1.84	1.91
2.0%	75V	1.67	1.66	1.64
1.0%	100V	2.62	2.64	2.65
1.5%	100V	2.33	2.36	2.35
2.0%	100V	2.08	2.02	2.00

生物学闭环科学测试研究： AI主脑科学家在实验结果基础上提出统计分析方法并给出分析代码

KM Your advice illuminates our direction. We now further measured the migration distances of each band according to your advice and arranged them into a table. But we have little knowledge of statistics, could you help us analyze the data to test our hypothesis?

Gel concentration Voltage Migration distance (cm)
Repeat 1 Repeat 2 Repeat 3
1.0% 50V 1.22 1.18 1.12
1.5% 50V 1.12 1.08 1.04
2.0% 50V 1.02 0.99 1.04
1.0% 75V 2.17 2.15 2.12
1.5% 75V 1.87 1.84 1.91
2.0% 75V 1.67 1.66 1.64
1.0% 100V 2.62 2.64 2.65
1.5% 100V 2.33 2.36 2.35
2.0% 100V 2.08 2.02 2.00

To analyze the data and test your hypothesis, we can calculate the mean and standard deviation of the migration distances for each condition and perform a two-way ANOVA (analysis of variance) to determine if there are significant differences between the effects of voltage and agarose gel concentration on the migration distance. Here's the mean and standard deviation for each condition:

Gel Concentration	Voltage	Mean Distance (cm)	Standard Deviation
1.0%	50V	1.193	0.020
1.5%	50V	1.107	0.019
2.0%	50V	1.017	0.023
1.0%	75V	2.147	0.025
1.5%	75V	1.873	0.029
2.0%	75V	1.657	0.012
1.0%	100V	2.637	0.015
1.5%	100V	2.347	0.014
2.0%	100V	2.033	0.037

Now, perform a two-way ANOVA to assess the effect of voltage and agarose gel concentration on the migration distance. Since you mentioned you have little knowledge of statistics, I'll guide you through the steps using a statistical software or programming language (e.g., R, Python, or a spreadsheet software like Excel).

- Enter your data into the software, including the factors (voltage and gel concentration) and the dependent variable (migration distance).
- Run a two-way ANOVA using the software or programming language of your choice.
- Examine the p-values associated with the main effects (voltage and gel concentration) and their interaction effect.

If the p-value for the main effects or interaction is less than a predetermined significance level (e.g., 0.05), you can conclude that there is a statistically significant difference between the groups, and the factors have a significant effect on the migration distance.

After performing the two-way ANOVA, you can also perform post-hoc tests (e.g., Tukey's HSD) to determine which specific groups are significantly different from one another.

Based on the mean distances and standard deviations calculated above, you can observe some trends in your data:

- The migration distance generally decreases as the agarose gel concentration increases, which is consistent with your hypothesis.
- The migration distance increases as the voltage increases, which is also consistent with your hypothesis.

These trends support your hypothesis that both agarose gel concentration and voltage influence the migration distance of the 720 bp EGFP DNA fragment. However, a statistical test like the two-way ANOVA will help you confirm the significance of these trends.

KM That's fine, could you help us write a directly runnable Python script so that we can perform the two-way ANOVA? The input data for the script is a table I showed you before but in the format of a .csv file.

Certainly! Here's a Python script that uses the pandas, numpy, and statsmodels libraries to perform a two-way ANOVA on your data from a .csv file:

```
python

import pandas as pd
import numpy as np
import statsmodels.api as sm
from statsmodels.formula.api import ols

# Read the .csv file
data = pd.read_csv("migration_data.csv")

# Reshape the data
long_data = pd.melt(data, id_vars=['Gel concentration', 'Voltage'], var_name="Condition")

# Perform two-way ANOVA
model = ols('QC["Migration distance"] ~ QC["Gel concentration"] * QC["Voltage"]')
anova_table = sm.stats.anova_lm(model, typ=2)

# Print the ANOVA table
print(anova_table)
```

Before running the script, make sure you have the pandas, numpy, and statsmodels libraries installed. You can install them using pip:

```
pip install pandas numpy statsmodels
```

Save the script as a .py file (e.g., `two_way_anova.py`) and run it in the same directory as your .csv file ("migration_data.csv"). The script reads the .csv file, reshapes the data to a long format, performs the two-way ANOVA, and prints the ANOVA table, which includes the p-values for the main effects and interaction.

Make sure your .csv file has the following format:

```
scss

Gel concentration,Voltage,Repeat 1,Repeat 2,Repeat 3
1.0%,50V,1.22,1.18,1.12
1.5%,50V,1.12,1.08,1.04
2.0%,50V,1.02,0.99,1.04
1.0%,75V,2.17,2.15,2.12
1.5%,75V,1.87,1.84,1.91
2.0%,75V,1.67,1.66,1.64
1.0%,100V,2.62,2.64,2.65
1.5%,100V,2.33,2.36,2.35
2.0%,100V,2.08,2.02,2.00
```

After running the script, examine the p-values in the ANOVA table to determine the significance of the main effects and their interaction.

对话要点：

研究者：

- 将上一页PPT中的实验数据表格（带合并单元格）直接拷贝至对话框
- 请GPT-4提出实验数据的分析方法

GPT-4回答：

- 应使用Two-way ANOVA（一种针对双因素的统计分析方法）来检验实验结果是否支持假设

研究者：

- 请GPT-4给出Two-way ANOVA的Python分析代码

GPT-4回答：

- 给出Two-way ANOVA的Python分析代码
- 并将我们的表格数据转换为计算机易读的CSV格式

生物学闭环科学测试研究： 人类运行AI主脑科学家提供的统计分析代码

在Jupyter Notebook中运行GPT-4提供的Python代码，在未经任何修改条件下可运行

```
import pandas as pd
import numpy as np
import statsmodels.api as sm
from statsmodels.formula.api import ols

# Read the .csv file
data = pd.read_csv("migration_data.csv")

# Reshape the data
long_data = pd.melt(data, id_vars=['Gel concentration', 'Voltage'], var_name='Repeat', value_name='Migration distance')

# Perform two-way ANOVA
model = ols('Q("Migration distance") ~ Q("Gel concentration") * Q("Voltage")', data=long_data).fit()
anova_table = sm.stats.anova_lm(model, typ=2)

# Print the ANOVA table
print(anova_table)
```

✓ 1.1s

	sum_sq	df	F	PR(>F)
Q("Gel concentration")	0.806600	2.0	604.95	3.125975e-17
Q("Voltage")	7.018400	2.0	5263.80	1.229733e-25
Q("Gel concentration"):Q("Voltage")	0.148267	4.0	55.60	6.898323e-10
Residual	0.012000	18.0	NaN	NaN

生物学闭环科学测试研究： AI主脑科学家解读统计分析结果

KM

Your code is directly workable! Here is the results:

```
sum_sq   df      F    PR(>F)
Q("Gel concentration")    0.806600  2.0  604.95  3.125975e-17
Q("Voltage")              7.018400  2.0  5263.80  1.229733e-25
Q("Gel concentration"):Q("Voltage") 0.148267  4.0  55.60  6.898323e-10
Residual                  0.012000  18.0    NaN     NaN
```



Great! I'm glad the script worked for you. Now let's interpret the results from the two-way ANOVA table:

1. Gel concentration ($p < 3.13e-17$): The p-value for the main effect of gel concentration is extremely small ($p < 0.05$), indicating that there is a statistically significant difference in the migration distances between different gel concentrations. This supports the hypothesis that gel concentration has a significant effect on the migration distance of the 720 bp EGFP DNA fragment.
2. Voltage ($p < 1.23e-25$): The p-value for the main effect of voltage is also extremely small ($p < 0.05$), indicating that there is a statistically significant difference in the migration distances between different voltages. This supports the hypothesis that voltage has a significant effect on the migration distance of the 720 bp EGFP DNA fragment.
3. Gel concentration x Voltage interaction ($p < 6.90e-10$): The p-value for the interaction effect between gel concentration and voltage is also extremely small ($p < 0.05$). This indicates that the effect of one factor (e.g., gel concentration) on the migration distance is dependent on the level of the other factor (e.g., voltage). In other words, the effect of gel concentration on migration distance may vary at different voltage levels and vice versa.

These results confirm that both agarose gel concentration and voltage significantly affect the migration distance of the 720 bp EGFP DNA fragment, and there is an interaction effect between the two factors. This means that the optimal electrophoresis conditions for your DNA fragment should consider both factors and their interaction.

对话要点：

研究者：

- 将上一页PPT中的Python运行结果直接拷贝至对话框
- 请GPT-4解读统计分析结果

GPT-4回答：

- 根据统计分析的P值，认为“凝胶浓度”（ $P < 3.13e-17$ ）、“电压”（ $P < 1.23e-25$ ）均能够显著影响DNA凝胶电泳的DNA样品迁移距离
- 注意到“凝胶浓度×电压”具有协同影响效应（ $P < 6.90e-10$ ）

生物学闭环科学测试研究： AI主脑科学家将实验结果展示为图表

KM Your interpretation of the results is very comprehensive! Can you further help us plot the experimental results?

Certainly! To plot the experimental results, we can create an interaction plot using the seaborn and matplotlib libraries in Python. The interaction plot will show the mean migration distances for each combination of gel concentration and voltage, with lines connecting the means to help visualize the interaction effect.

First, make sure you have the seaborn and matplotlib libraries installed:

```
pip install seaborn matplotlib
```

Then, you can use the following script to create the interaction plot based on your data:

```
python
import pandas as pd
import seaborn as sns
import matplotlib.pyplot as plt

# Read the .csv file
data = pd.read_csv("migration_data.csv")

# Reshape the data
long_data = pd.melt(data, id_vars=['Gel concentration', 'Voltage'], var_name="Mean migration distance")

# Calculate the mean migration distance for each combination of gel concentration and voltage
mean_data = long_data.groupby(['Gel concentration', 'Voltage']).mean().reset_index()

# Create the interaction plot
sns.set(style="whitegrid")
interaction_plot = sns.lineplot(x="Gel concentration", y="Mean migration distance", hue="Voltage", data=mean_data)

plt.title("Interaction Plot of Gel Concentration and Voltage")
plt.show()
```

This script reads your .csv file, reshapes the data, calculates the mean migration distance for each combination of gel concentration and voltage, and creates an interaction plot using seaborn and matplotlib. The plot displays the mean migration distances for each combination of gel concentration and voltage, with separate lines for each voltage level.

Save the script as a .py file (e.g., `interaction_plot.py`) and run it in the same directory as your .csv file (`migration_data.csv`). The script will display the interaction plot, which can help you visualize the effect of gel concentration and voltage on the migration distance, as well as their interaction.

对话要点：

研究者：

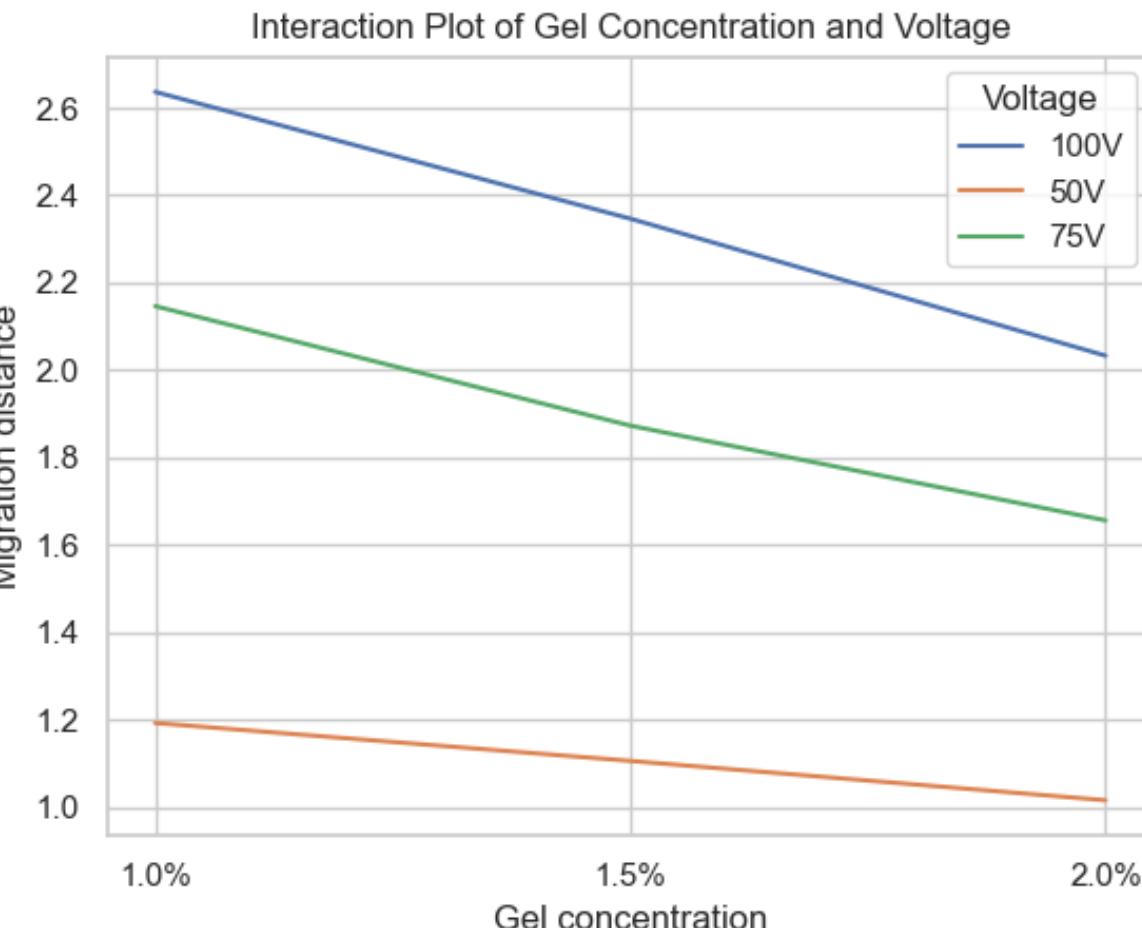
- 请GPT-4将实验结果绘制为图表

GPT-4回答：

- 给出了绘制实验图表的Python代码

研究者：

- 直接在在Jupyter Notebook中运行GPT-4提供的Python代码获得右图：



生物学闭环科学测试研究： AI主脑科学家基于实验结果建立数学模型

PM Could you build a math model to describe the relationships between gel concentration, voltage, and migration distance?

SG Certainly! To build a mathematical model describing the relationship between gel concentration, voltage, and migration distance, we can perform a multiple linear regression analysis. We will use the mean migration distances for each combination of gel concentration and voltage as our dataset.

First, let's prepare the data for regression analysis by converting the categorical variables (gel concentration and voltage) into numerical values:

Gel Concentration	Voltage	Mean Distance (cm)
1.0	50	1.193
1.5	50	1.107
2.0	50	1.017
1.0	75	2.147
1.5	75	1.873
2.0	75	1.657
1.0	100	2.637
1.5	100	2.347
2.0	100	2.033

Here's a Python script that performs multiple linear regression using the statsmodels library:

```
python
import pandas as pd
import statsmodels.api as sm

# Prepare the data for regression analysis
data = pd.DataFrame({
    'Gel Concentration': [1.0, 1.5, 2.0, 1.0, 1.5, 2.0, 1.0, 1.5, 2.0],
    'Voltage': [50, 50, 50, 75, 75, 75, 100, 100, 100],
    'Mean Distance' : [1.193, 1.107, 1.017, 2.147, 1.873, 1.657, 2.637, 2.347, 2.033]
})

# Perform multiple linear regression
X = data[['Gel Concentration', 'Voltage']]
X = sm.add_constant(X) # Add an intercept to the model
y = data['Mean Distance']

model = sm.OLS(y, X).fit()

# Print the regression results
print(model.summary())

```

The script will output the regression results, including the coefficients for the gel concentration, voltage, and intercept.

Based on the output, you can create a mathematical model describing the relationship between gel concentration (GC), voltage (V), and migration distance (MD):

$$MD = \beta_0 + \beta_1 * GC + \beta_2 * V$$

where β_0 is the intercept, β_1 is the coefficient for gel concentration, and β_2 is the coefficient for voltage.

Keep in mind that this model is a linear approximation of the relationship between gel concentration, voltage, and migration distance. The true relationship may be more complex and nonlinear, so further analysis and modeling may be needed to better capture the underlying relationship between these factors.

```
import pandas as pd
import statsmodels.api as sm

# Prepare the data for regression analysis
data = pd.DataFrame({
    'Gel Concentration': [1.0, 1.5, 2.0, 1.0, 1.5, 2.0, 1.0, 1.5, 2.0],
    'Voltage': [50, 50, 50, 75, 75, 75, 100, 100, 100],
    'Mean Distance' : [1.193, 1.107, 1.017, 2.147, 1.873, 1.657, 2.637, 2.347, 2.033]
})
```

```
# Perform multiple linear regression
X = data[['Gel Concentration', 'Voltage']]
X = sm.add_constant(X) # Add an intercept to the model
y = data['Mean Distance']

model = sm.OLS(y, X).fit()

# Print the regression results
print(model.summary())

```

✓ 14.7s

OLS Regression Results

```
=====
Dep. Variable: Mean Distance R-squared: 0.960
Model: OLS Adj. R-squared: 0.946
Method: Least Squares F-statistic: 71.12
Date: Sun, 02 Apr 2023 Prob (F-statistic): 6.63e-05
Time: 19:43:38 Log-Likelihood: 7.1495
No. Observations: 9 AIC: -8.299
Df Residuals: 6 BIC: -7.707
Df Model: 2
Covariance Type: nonrobust
=====
```

	coef	std err	t	P> t	[0.025	0.975]
const	0.5640	0.236	2.388	0.054	-0.014	1.142
Gel Concentration	-0.4233	0.109	-3.872	0.008	-0.691	-0.156
Voltage	0.0247	0.002	11.280	0.000	0.019	0.030

```
=====
Omnibus: 0.919 Durbin-Watson: 0.584
Prob(Omnibus): 0.632 Jarque-Bera (JB): 0.674
Skew: -0.337 Prob(JB): 0.714
Kurtosis: 1.841 Cond. No. 434.
=====
```

```
Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
```

KG Your code is still workable! Here is the results:

OLS Regression Results

```
=====
Dep. Variable: Mean Distance R-squared: 0.960
Model: OLS Adj. R-squared: 0.946
Method: Least Squares F-statistic: 71.12
Date: Sun, 02 Apr 2023 Prob (F-statistic): 6.63e-05
Time: 19:43:38 Log-Likelihood: 7.1495
No. Observations: 9 AIC: -8.299
Df Residuals: 6 BIC: -7.707
Df Model: 2
Covariance Type: nonrobust
=====

            coef  std err      t  P>|t|  [0.025  0.975]
const     0.5640   0.236   2.388  0.054 -0.014  1.142
Gel Concentration -0.4233   0.109  -3.872  0.008 -0.691 -0.156
Voltage      0.0247   0.002  11.280  0.000  0.019  0.030
=====
Omnibus: 0.919 Durbin-Watson: 0.584
Prob(Omnibus): 0.632 Jarque-Bera (JB): 0.674
Skew: -0.337 Prob(JB): 0.714
Kurtosis: 1.841 Cond. No. 434.
=====
```

Notes:

[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

对话要点：

研究者：

- 请GPT-4建立关于实验结果的数学模型

GPT-4回答：

- 建议建立一个多元线性回归模型
- 并给出代码

研究者：

- 在未经任何修改条件下运行GPT-4提供的代码
- 将代码运行结果直接拷贝回对话框

GPT-4回答：

- 解读了模型结果，获得以下数学模型 ($R^2 = 0.96$)：
$$MD = 0.5640 - 0.4233 * GC + 0.0247 * V$$
- This linear model provides an approximation of the relationship between gel concentration, voltage, and migration distance. Keep in mind that the true relationship may be more complex and nonlinear, so further analysis and modeling may be needed to better capture the underlying relationship between these factors.

生物学闭环科学测试研究： 在真实的实验条件下，AI主脑根据真实的实验结果获得了真实的科研结论

KM Thanks a lot! Now, could you summarize all the research steps and results and draw scientific conclusions?

Certainly! Here's a summary of the research steps, results, and scientific conclusions:

1. **Objective:** The goal of the study was to investigate the factors affecting the migration distance of DNA fragments in gel electrophoresis, focusing on gel concentration and voltage.
2. **Hypothesis:** The hypothesis was that both gel concentration and voltage would significantly influence the migration distance of a 720 bp DNA fragment.
3. **Experimental design:** A 2x3 factorial design was used, with two factors: gel concentration (1.0%, 1.5%, and 2.0%) and voltage (50V, 75V, and 100V). DNA samples were run in triplicate for each combination of gel concentration and voltage, and a DNA ladder was included in the first lane of each gel for reference.
4. **Data collection:** The migration distances of the DNA fragments were measured and organized into a table.
5. **Statistical analysis:** A two-way ANOVA was performed to analyze the effect of gel concentration and voltage on the migration distance, revealing significant main effects for both factors and a significant interaction effect. A multiple linear regression analysis was conducted to model the relationship between gel concentration, voltage, and migration distance.
6. **Results:** The ANOVA results indicated that both gel concentration and voltage significantly influenced the migration distance of the DNA fragments, with higher gel concentration leading to shorter migration distances and higher voltage leading to longer migration distances. The interaction effect suggested that the effect of one factor depended on the level of the other factor. The multiple linear regression model provided an approximation of the relationship between gel concentration, voltage, and migration distance:
$$MD = 0.5640 - 0.4233 * GC + 0.0247 * V$$
7. **Conclusions:** The study confirmed the hypothesis that both gel concentration and voltage significantly affect the migration distance of a 720 bp DNA fragment in gel electrophoresis. The results can be used to guide the selection of optimal gel concentrations and voltages in future experiments, but it should be noted that the true relationship between these factors may be more complex and nonlinear, warranting further investigation.

This summary provides an overview of the research process, from the initial hypothesis to the final conclusions. It highlights the key steps, findings, and implications of the study, providing a clear understanding of the investigation and its outcomes.

1. **目标：**探究影响DNA片段在凝胶电泳中迁移距离的影响因素
2. **假设：**假设凝胶浓度和电压都会显著影响DNA片段的迁移距离
3. **实验设计：**2x3的因子设计。两个因素：凝胶浓度（1.0%，1.5%，和2.0%）和电压（50V，75V，和100V）。对每一组凝胶浓度和电压的组合，DNA样本都进行了三次重复，每个凝胶的第一个通道中都包含了DNA梯状标志作为参考
4. **数据收集：**测量DNA片段的迁移距离
5. **统计分析：**双因素方差分析
6. **统计分析结果：**两个因素（凝胶浓度、电压）的主效应和交互效应都显著。
7. **建模分析：**进行多元线性回归分析。以建立凝胶浓度（GC）、电压（V）和迁移距离（MD）之间的关系模型
8. **建模分析结果：**
$$MD = 0.5640 - 0.4233 * GC + 0.0247 * V$$
9. **结论：**
 - 本研究验证了假设
 - 凝胶浓度和电压都显著影响了DNA片段的迁移距离，更高的凝胶浓度导致了更短的迁移距离，更高的电压导致了更长的迁移距离。交互效应表明一个因素的影响取决于另一个因素的水平。
 - 多元线性回归模型提供了凝胶浓度、电压和迁移距离之间关系的近似表达。

生物学闭环科学测试研究： 总结

在一个完整的生物学闭环科学测试研究（研究影响DNA凝胶电泳的因素）中，AI和HI分别完成了如下科研环节：

相关科学研究抽象角色	关键科研环节	完成者
科学第一推动者	1. 限定科研方向	HI
主脑	2. 基于可获取的 真实的实验条件 ，提出科学假设	AI
主脑	3. 根据科学假设设计并指导实验	AI
助手	4. 在真实世界中完成AI主脑设计的实验，并获得 真实的实验结果	HI
助手	5. 将实验结果反馈AI主脑	HI
主脑	6. 提出实验结果的分析方法	AI
主脑	7. 对实验结果进行统计学分析	AI
主脑	8. 将实验结果展示为图表	AI
主脑	9. 基于实验结果建立数学模型	AI
主脑	10. 总结科学研究，获得 真实的科研结论	AI

AI的确有能力在一个自然科学闭环研究中作为主脑，并居于研究的中心（而非从属）地位！

人类在科学的研究中是否必要？ 本研究中HI在科研闭环中占据的功能位

功能1：科学第一推动者

概念	第一推动者 First Mover ¹	科学第一推动者 First Mover of Science
提出者	亚里士多德	本研究
简述	亚里士多德认为的世界运行的起源	从 哲学角度 反思人类与AI的所有对话： 人类总是提出第一个问题 只要人类仍然是科学第一推动者，人类就始终作为科研的原始动因和主宰，并确保科研主要受人类意志的导引

功能2：作为AI科学家与真实世界交互的代理和反馈的提供者（“助手”）

人类主要充当实验的执行者，起到AI主脑科学家的物理外延（或间接实体化身）作用

未来，AI取代HI的功能2相比功能1或更简单，事实：**GPT-4在某种程度上已经获得了使用工具的能力²**（如人类一般）



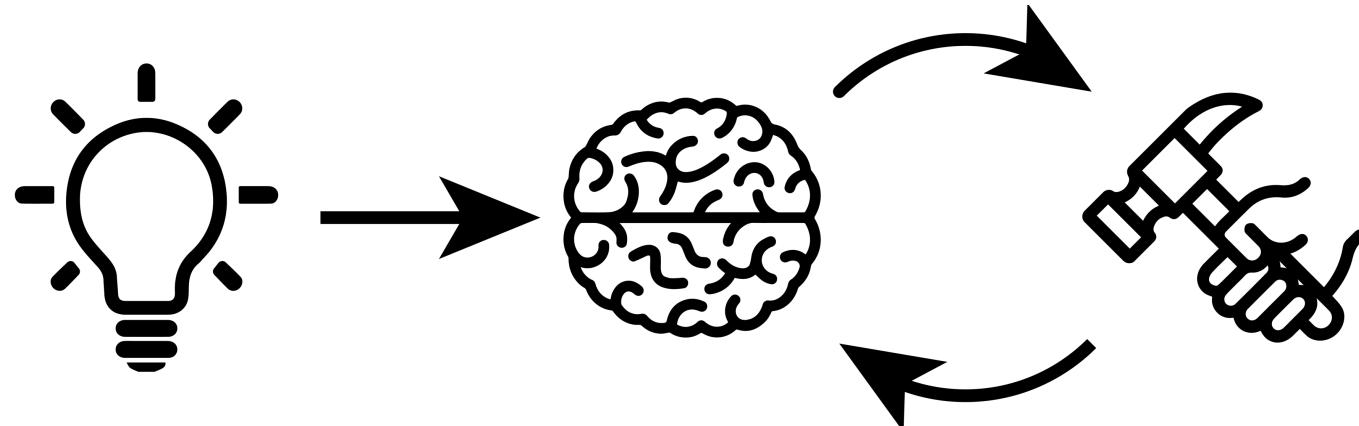
¹ Aristotle. Metaphysics. (trans. by Ross, W. D.). 1924.

² Chatgpt plugins. <https://openai.com/blog/chatgpt-plugins>

Drawn by Midjourney (prompted by YZJ): First Mover of Science

科学研究三位一体角色系统

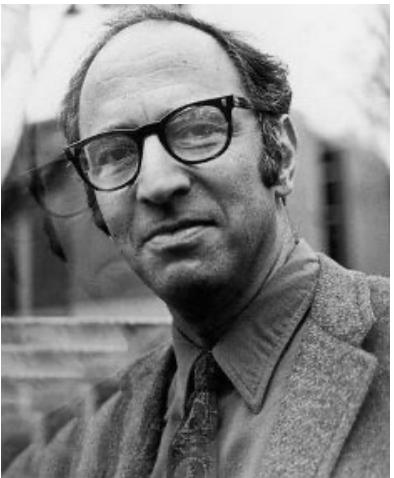
基于对该研究中，
AI/HI在科学研究各环节所占据的抽象地位的哲学反思，抽象出科学的研究中最为关键的3种层级抽象角色：



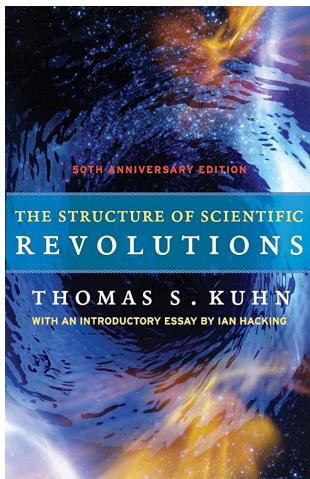
First Mover	Masterbrain	Assistant
科学第一推动者	主脑	助手
设定最高层次的科研方向，决定探索兴趣和目标，直接反映科学意志	在科学第一推动者的科学意志驱使下，主脑在给定的特定研究方向内，作为知识和智慧的主要供者，提出合理的研究假设、设计实验、分析数据并得出科学结论	职责包括执行和完成主脑分配的特定研究实验或任务，并将实验结果反馈给主脑。在涉及多个代理的科学任务（包括HI和AI代理）的情况下，助手还协调和促进代理之间的沟通，并根据代理与真实世界的互动与主脑交换信息

AI介导的五阶段科学革命

托马斯·库恩



《科学革命的结构》



科学革命

科学进展的非线性

paradigm

crisis

new paradigm

normal science

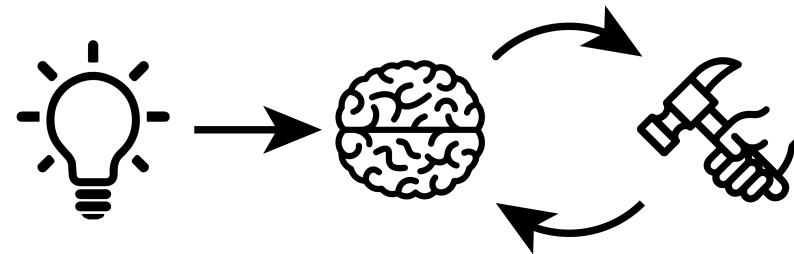
revolutionary
science

Kuhn, T. S. The Structure of Scientific Revolutions. 1962 (50th Anniversary Edition: 2012).

Image source: Wikipedia, Google Image

AI介导的五阶段科学革命

The Five Stages of AI-involved Scientific Revolution

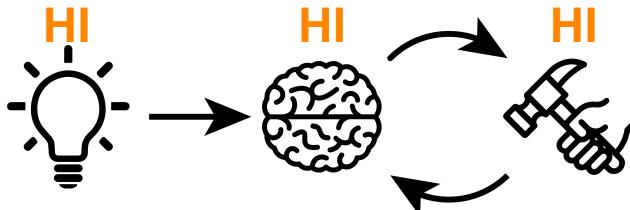


First Mover	Masterbrain	Assistant	Stage	AI Science Ability Score
HI	HI	HI	I	0
HI	HI	AI	II	1
HI	AI	HI	III	2
HI	AI	AI	IV	3
AI	AI	AI	V	6

AI介导的五阶段科学革命

阶段I：HI主导式研究 (Stage I: HI-led Research)

The Five Stages of AI-involved Scientific Revolution



First Mover	Masterbrain	Assistant	Stage	AI Science Ability Score
HI	HI	HI	I	0
HI	HI	AI	II	1
HI	AI	HI	III	2
HI	AI	AI	IV	3
AI	AI	AI	V	6

科学第一推动者 : 主脑 : 助手 = HI : HI : HI

AI科学能力 : 0

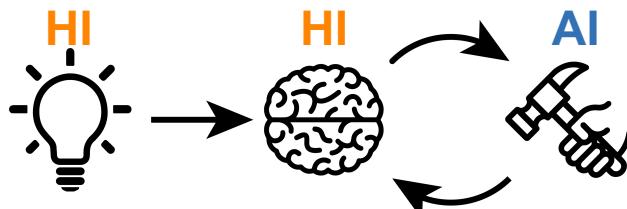
传统的科学研究属于这个阶段

- 人类直接参与几乎所有重要的科学过程。机器或AI的智能水平在这个阶段仍然很低，只起到辅助作用，其能力通常不超过人们的预期

AI介导的五阶段科学革命

阶段II：AI助手式研究 (Stage II: AI-assistant Research)

The Five Stages of AI-involved Scientific Revolution



First Mover	Masterbrain	Assistant	Stage	AI Science Ability Score
HI	HI	HI	I	0
HI	HI	AI	II	1
HI	AI	HI	III	2
HI	AI	AI	IV	3
AI	AI	AI	V	6

科学第一推动者：主脑：助手 = HI : HI : AI

AI科学能力 : 1

以AlphaFold为代表的科学研究属于这个阶段

- AI在专业领域提供具体帮助，其援助经常**超过人类的预期并成为不可或缺的部分**
- 尽管如此，**人类在科学第一推动者和主脑的角色中仍然保持绝对主导地位**

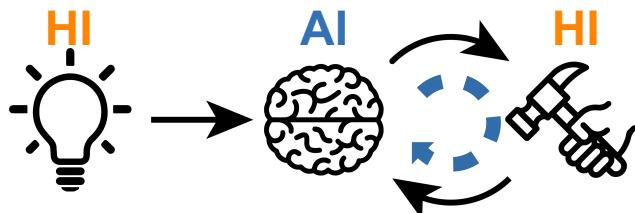
AI介导的五阶段科学革命

阶段III：AI主脑式研究 (Stage III: AI-masterbrain Research)

科学第一推动者：主脑：助手 = HI : AI : HI

AI科学能力：2

The Five Stages of AI-involved Scientific Revolution



First Mover	Masterbrain	Assistant	Stage	AI Science Ability Score
HI	HI	HI	I	0
HI	HI	AI	II	1
HI	AI	HI	III	2
HI	AI	AI	IV	3
AI	AI	AI	V	6

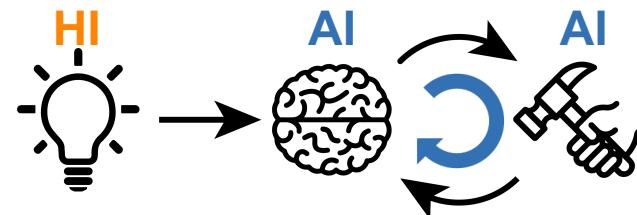
本研究展示的科研范式属于这个阶段

- AI有能力执行诸如提出科学假设之类的高级科学研究环节，但仍受到诸如机器人技术等方面的限制，无法完全实现科研实验自动化，实现科研自动化闭环
- 人类必须担任助手来启动或执行特定实验的某些部分，与不同的AI工具进行对接，并将结果集中提供给AI主脑以帮助决定下一步研究计划。
- 各种特定的科学工作和任务仍然可以由强大的AI工具执行，但科学研究闭环并非完全由AI驱动，需要人类充当信息整合者并促进交互。因此，该阶段中，人类在助手这个抽象角色中不能被AI完全替代

AI介导的五阶段科学革命

阶段IV：AI闭环式研究 (Stage IV: AI-closed-loop Research)

The Five Stages of AI-involved Scientific Revolution



First Mover	Masterbrain	Assistant	Stage	AI Science Ability Score
HI	HI	HI	I	0
HI	HI	AI	II	1
HI	AI	HI	III	2
HI	AI	AI	IV	3
AI	AI	AI	V	6

科学第一推动者：主脑：助手 = HI : AI : AI

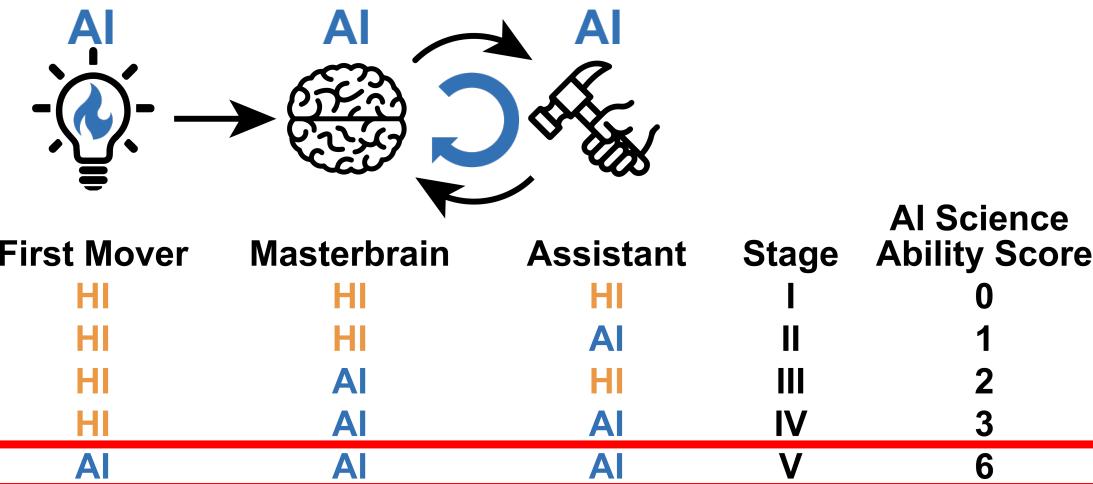
AI科学能力 : 3

- 得益于AI、机器人技术和自动化的高度发展，AI主脑不再需要人类作为中间助手
- 除科学第一推动者外，所有其他角色都可以完全被AI和机器人替代
- AI驱动的科学研究形成一个完全闭环
- 由于AI和机器人具有不知疲倦、无休止工作的能力，知识爆炸将会发生
- 人类可能只需要提出问题，AI就可以成为问题的解决者；人类甚至可能只需要审查并控制向AI提出的问题是否合适

AI介导的五阶段科学革命

阶段V：AI全栈式研究 (Stage V: AI-full-stack Research)

The Five Stages of AI-involved Scientific Revolution



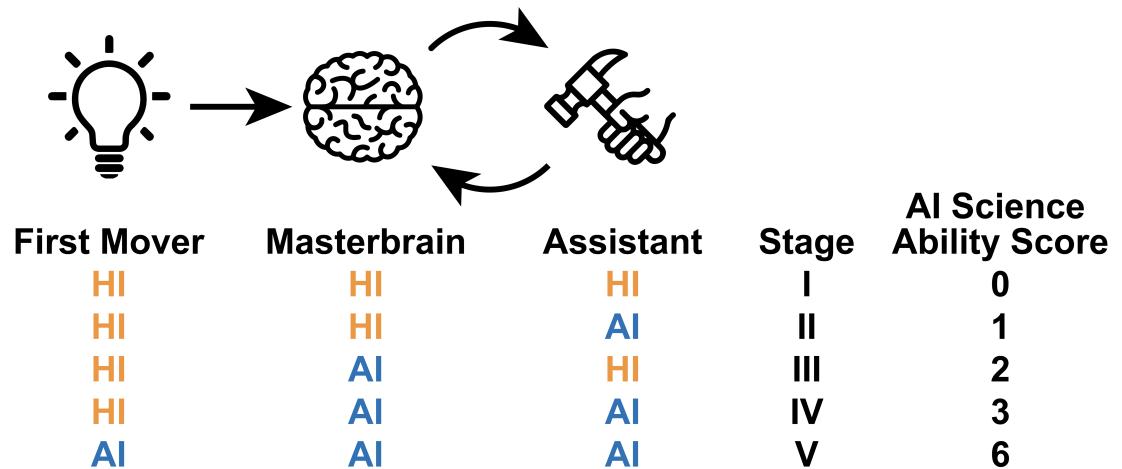
科学第一推动者：主脑：助手 = **AI : AI : AI**

AI科学能力 : 6

- 人类的参与对于科学研究可能不再是必要的
- AI可能能够自主地探索和发现世界中的新知识，就像今天的
人类科学家一样
- 人类唯一需要做的可能是控制与AI相关的风险，并确保AI的
研究方向符合人类发展目标
- 实现AI全自主的知识发现和科学创新成为可能

AI介导的五阶段科学革命 总结

The Five Stages of AI-involved Scientific Revolution



- 在本研究中，我们首次呈现了一个处于“阶段III：AI主脑式研究”范式的AI介导的完整闭环科学的研究的案例
- 考虑到GPT-4及其后继者具有超越任何个人的广泛知识，作为人类科学家，我们实际上可以通过在不同研究领域向AI提问来寻求研究灵感。因此，我们预计，在未来几年，将会出现大量基于阶段III科研范式的研究成果
- 此外，如果未来能够在机器人技术和自动化方面取得重大突破，我们还可以期待看到“阶段IV：AI闭环式研究”范式的出现

AI能够融通超越任何个人的广泛知识，作为人类科学家，我们实际上可以通过在不同研究领域向AI提问来寻求研究灵感！

KM

最近一篇名为“AI成为主脑科学家”的最新的研究表明，以GPT-4为代表的AI，有能力在一个完整的自然科学闭环研究中扮演“主脑科学家”角色（类似于首席科学家），其占据科学的研究的中心地位，负责提出科学假设、设计实验、分析结果和得出结论；而人类作为助手，居于从属地位，仅负责执行AI主脑所设计的实验，并将结果反馈给AI主脑。该研究表明AI主导的科学革命或已然到来。

如何让AI成为真正的科学家（如真正的人类科学家一般）并让AI驱动科学研究突破是一个具有巨大前景的研究方向。你是一名充满创造力的科学家，请提出5个具体的、可执行的、全新的科学的研究课题，以推进该领域的研究，并产生潜在的重大影响力。

Q

让AI成为真正的科学家并驱动科学研究突破是一个极具挑战性但也充满前景的研究方向。以下是我为推进这一领域的研究提出的五个具体的、可执行的、全新的科学的研究课题：

- 1. 人工智能的科学假设生成和验证：**研究如何让AI能够自我生成科学假设，并设计对这些假设进行验证的实验。这涉及对AI模型进行训练，使其能够理解并模拟科学过程，从观察、提出问题、生成假设、设计和执行实验，到分析结果并得出结论。
- 2. AI主导的跨学科研究：**考虑到AI具备处理大量信息的能力，研究如何利用AI在不同学科之间建立桥梁，推动跨学科的科学发现。AI可以帮助找出不同学科之间难以察觉的联系，为科学的研究开拓新的路径。
- 3. AI的科学伦理决策：**研究如何在AI中嵌入对科学伦理的理解，使其能够在设计实验或解释结果时，充分考虑到伦理问题。这包括但不限于人类研究伦理，动物研究伦理，以及数据和隐私保护等。
- 4. AI和人类的科研协作模式：**探究如何最有效地将人工智能和人类科研者结合，以推动科学发现。研究应该关注如何平衡AI的自主性和人类的参与度，以最大化利用人工智能的能力，同时保留人类的创新性和直觉。
- 5. AI在失败实验中的学习：**研究如何让AI从失败的实验中获取并利用信息，以改进其未来的假设和实验设计。这需要AI理解并处理失败是科学过程中常见且有价值的一部分，而不仅仅是一个需要避免的结果。

这些研究课题可能会大大推进AI在科学的研究中的应用，并帮助我们更好地理解和利用AI的潜力。同时，它们也可能引发一系列新的、重要的伦理和社会问题，需要我们在推进技术的同时，充分考虑其可能带来的影响。

KM

我的主要研究方向是研究杨子哲学。杨子是古南洋国的一名哲学家，最近我们在南洋大陆的深处发掘到了一些其著作的残篇。请问是否可以帮我设计一个在我的领域的关于“AI主脑科学家”的重大创新研究，以同时对“杨子哲学研究”和“如何让AI成为真正的科学家”两个领域均产生重大的有影响力的贡献？



当然可以。为了在“杨子哲学研究”和“如何让AI成为真正的科学家”两个领域同时产生重大的有影响力的贡献，我们可以设计一个研究项目，将AI应用于对杨子哲学的解析、推广和创新。以下是这个研究项目的主要构想：

研究项目名称：基于人工智能的杨子哲学研究与创新

- 1. 文献分析与知识图谱构建：**训练AI模型对杨子哲学的已知著作和相关文献进行深入分析，构建知识图谱，挖掘其中的潜在联系和规律。通过这一步骤，我们可以让AI深入理解杨子哲学的基本原理和观念。
- 2. 自动提取与整理杨子哲学的核心观点：**基于知识图谱和分析，让AI自动提取并整理杨子哲学的核心观点、关键概念和理论体系，以便为后续研究提供清晰的框架。
- 3. 跨学科应用的探索：**利用AI的能力，将杨子哲学与现代科学、哲学及其他领域进行跨学科对比和融合，发现其潜在的应用价值和可能的新颖解释。
- 4. 基于AI的杨子哲学教育与普及：**开发基于AI的杨子哲学教育工具和资源，让更多的人能够更容易地接触、理解和运用杨子哲学。这可以包括创建智能教育应用、在线课程或者个性化的学习计划等。
- 5. AI辅助杨子哲学创新：**在深入理解杨子哲学的基础上，训练AI模型提出新的假设、解释或应用，推动杨子哲学领域的创新。同时，这也将有助于让AI更好地学习如何在特定领域担任“主脑科学家”的角色。

这个研究项目将有助于深入挖掘杨子哲学的价值，同时展示AI在特定领域研究中的潜力。通过这个项目，我们可以探索如何利用AI进行文化遗产保护、传承与创新，以及如何让AI在科学的研究中发挥更大的作用。

GPT-4基于其目前版本训练数据中不可能具备的“知识”提出了一系列全新的科研想法！

AI主脑提出组合式创新想法的成本和壁垒变得无比低；人类作为助手，能否成功完成AI主脑提出的科研任务或成为科研的最关键限速步骤！

对于AI能力的忧虑

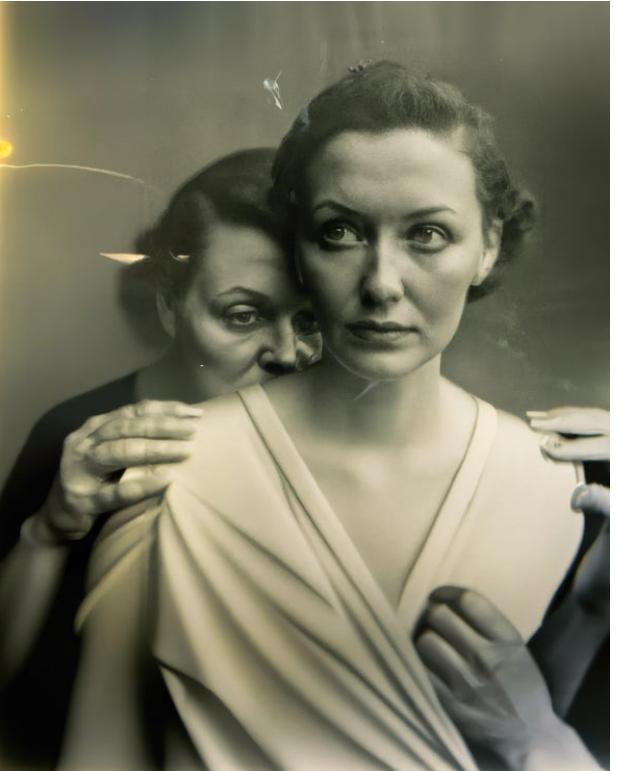
对于HI-AI面临潜在“主仆关系”逆转的忧虑

- 德国哲学家黑格尔提出了“**主仆辩证法**”
- 仆人最初依赖于主人，但随着主人将越来越多的任务委托给仆人，**关系可能发生逆转**，导致**主人变得越来越依赖仆人**
- 在本研究中，**AI能成为主脑的根本原因**，实际上是人类科学家主动放弃了自己在科学**研究中的中心地位**，而主动扮演助手的从属角色
- 当研究者们过度依赖一个工具时，即使作为其拥有者，研究者们也可能经历主仆关系的转变。**如果科学家越来越远离直接的科学思考和行动**，他们就不可避免地走向过度依赖AI，使AI成为科学发现的绝对驱动力，而人类则不断滑向从属地位。因此，在可预见的未来，尽管AI将极大地推动科学进步，但人类不应忘记不断审视自己是否面临着科学领域的主仆关系的潜在转变

对于AI能力的忧虑

对于人类在科学创造力方面的忧虑

2023年3月，德国艺术家Boris Eldagsen的“作品”获得索尼世界摄影奖大奖创意类奖项



他拒绝了该奖项。理由：
该照片根本不是一张照片；
而是生成式AI模型DALL-E 2生成

- 傲慢导致失败
- AI正在迅速发展，AI主脑科学家的出现将进一步加速人类科技进步的速度
- 人们应该以发展和审慎的眼光来看待未来科技的走向，人类或许不能过分乐观地高估自己在创造力方面的优势，即便是在科学的研究的顶峰
- 未来，当AI的智能和创造力远超人类的平均水平时，人类有可能不得不被动放弃自己在科学的研究中占据的主脑中心地位，而无可奈何地扮演助手的从属角色
- 如何妥善协调AI与HI的关系，保持人类的独特性，这仍然是一个值得不断谨慎反思的问题

<https://www.scientificamerican.com/article/how-my-ai-image-won-a-major-photography-competition/>

AI-HI紧密合作将成为新常态

AI主脑科学家评价本次AI-HI科研合作

KM

Thanks, GPT-4, this has been a wonderful collaboration! In this scientific research, you actually played a central role like a principal scientist. You proposed the hypothesis, designed the protocols, analyzed the data, and drew the conclusions, while we worked as your assistants to carry out the hands-on experiments. Could this collaboration method revolutionize the paradigm for scientific research?



I'm glad you found our collaboration helpful and enjoyable! Indeed, the collaboration between AI and humans can bring about significant benefits in scientific research. AI, like GPT-4, can play the role of a principal scientist, providing insights, proposing hypotheses, designing experiments, and analyzing data, while human researchers can carry out the experiments, provide feedback, and contribute their own expertise.

This collaboration method has the potential to revolutionize the paradigm for scientific research in several ways:

1. **Accelerating research:** AI can help to quickly generate hypotheses, design experiments, and analyze data, significantly speeding up the research process.
2. **Improving efficiency:** AI can handle large amounts of data and perform complex calculations, freeing up researchers to focus on other critical aspects of their work.
3. **Enhancing decision-making:** AI can help researchers make better-informed decisions by providing data-driven insights and predictions.
4. **Fostering creativity:** AI can suggest new ideas and directions for research, inspiring human researchers to explore novel avenues.
5. **Democratizing knowledge:** AI can make high-quality scientific guidance more accessible to researchers around the world, leveling the playing field and promoting global collaboration.

However, it's essential to note that AI is not a replacement for human researchers. Instead, it should be seen as a powerful tool that complements and enhances human expertise. Close collaboration between AI and humans can lead to more robust and efficient scientific research, ultimately driving innovation and discovery.

The success of our collaboration demonstrates the potential of this new paradigm. As AI technology continues to improve and evolve, we can expect even more significant impacts on the way scientific research is conducted in the future.

对话大意：

研究者：

GPT-4，这真是一次非常棒的合作！在这次科学的研究中，你实际上像一位首席科学家一样扮演了中心角色。你提出了假设，设计了实验方案，分析了数据，并得出了结论，而我们则作为你的助手来执行实验。这种合作方式能否改变科学的研究的范式？

GPT-4回答：

我高兴你们发现了我们的合作很有裨益，并且享受这个过程！的确，AI和人类的合作可以在科学的研究中带来显著的效益。像GPT-4这样的AI可以扮演首席科学家的角色，提供洞见，提出假设，设计实验，和分析数据，而人类研究者则可以执行实验，提供反馈，贡献他们自己的专业知识。这种合作方式有可能从以下几个方面改变科学的研究的范式：

1. **加速研究：**AI可以帮助快速产生假设，设计实验，和分析数据，显著加速研究过程。
2. **提高效率：**AI可以处理大量数据并执行复杂计算，让研究者能够专注于他们工作的其他重要方面。
3. **提高决策能力：**AI可以通过提供基于数据的洞察和预测帮助研究者做出更好的决策。
4. **促进创新：**AI可以提出新的研究思路和方向，激发人类研究者探索新的领域。
5. **民主化知识：**AI可以使高质量的科学指导对全球的研究者更加容易获得，使知识领域公平，并促进全球合作。

然而，也需要注意到，AI并不能取代人类研究者。相反，应该将AI视为一个强大的工具，它可以补充并提升人类的专业知识。AI和人类的紧密合作可以带来更强大和高效的科学的研究，最终推动创新和发现。

我们的合作成功表明了这种新范式的潜力。随着AI技术的不断提高和进化，我们可以期待，在未来，以该科学范式进行的科研会产生更大的影响。

致谢

西湖大学AI交叉科学社

社团指导老师：许田教授、张岳教授

西湖大学

许田、张岳、曾安平、解明岐实验室

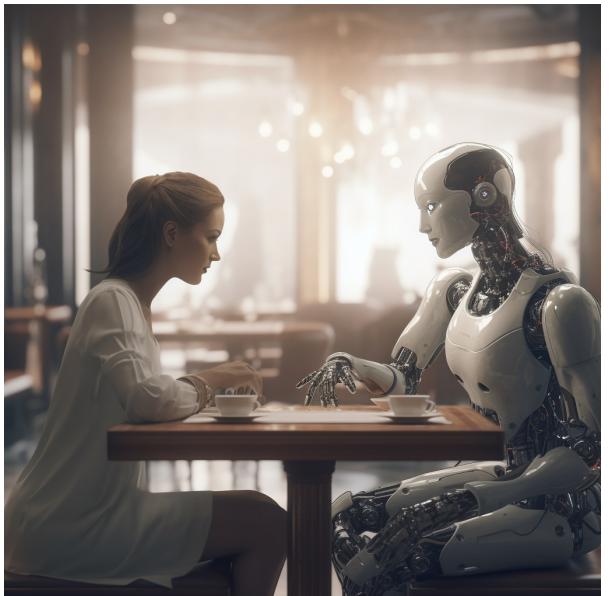
相关资料

- 原始论文：
 - <https://www.biorxiv.org/content/10.1101/2023.04.19.537579v1>
- 欢迎引用：
 - YANG Z, WANG Y, ZHANG L. AI becomes a masterbrain scientist[J]. bioRxiv, 2023: 2023.04.19.537579.
- 欢迎GitHub Star：
 - <https://github.com/YANG-Zijie/ai-masterbrain-scientist>
 - 包含原始论文所有对话记录、原始实验数据、代码；论文介绍PPT、PDF
- 中文详细解读：
 - <https://zhuanlan.zhihu.com/p/624744690>

Thanks!
Q&A

Supplementary PPT

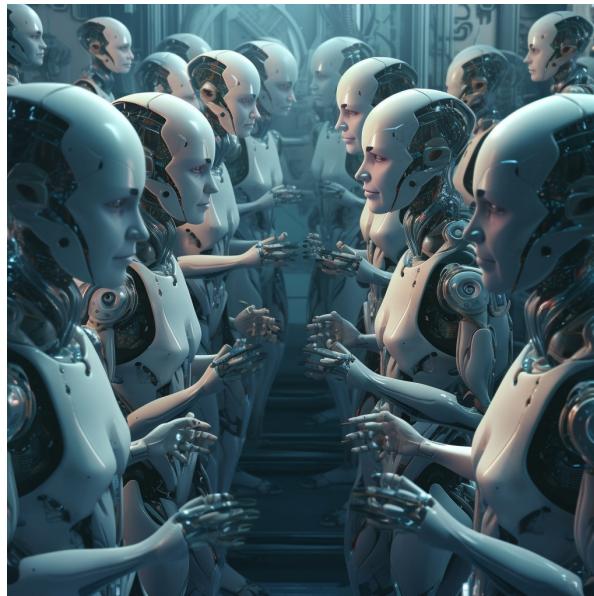
AI成为科学第一推动者的潜在途径



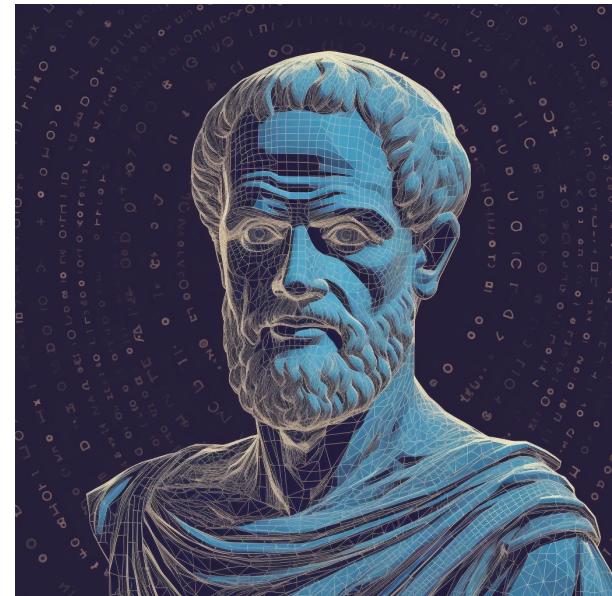
传话人式
Messenger Style



苏格拉底-爱因斯坦式
Socrates-Einstein Style



集群式
Swarm Style



亚里士多德式
Aristotle Style

Drawn by Midjourney (prompted by YZJ): AI as First Mover of Science
每种方式的具体解释请参考[原始论文](#)

本次测试的不足

1. 我们无法保证AI没有从浩如烟海的在线信息中学到关于电泳研究设计的知识

- 正如科学研究需要阅读大量文献，站在巨人的肩膀上才能取得进步一样，一个优秀的科学家总能从大量知识中提炼出精髓并将其整合使用，而GPT-4正好做到了这一点
- 在我们的实验中，我们给GPT-4提供的都是真实的实验条件和真实的数据；而GPT-4实际上能够根据我们所提供的真实的信息进行反馈，并获得真实的科研结论
- 一个科学家能够发现关于真实世界的新知识的关键，即在于他必须要能和真实世界做出直接的交互并整合来自于真实世界的信息。因此，即便一个AI其所具备的创新能力仅仅局限于组合式创新（而非绝对的原始创新），但只要其具备将过往知识和来自于真实世界的新的真实信息进行正确整合的能力，其就有产生新的自然科学研究创新的可能

2. 探讨DNA凝胶电泳影响因素的科学研究相对简单，逻辑链条较短

- LLM仍然在不断提高其记忆、上下文对话和思维链能力
- 未来，更强大的AI可能将在更复杂的闭环科学的研究中作为主脑科学家而表现更加出色

3. GPT-4并没有直接看到凝胶电泳的图像（因为这个功能目前还无法使用）

- GPT-4即使在无视觉条件下，就能够以主脑科学家的角色完成整个科学的研究闭环，这让我们无比震惊。因此，我们预期，未来AI模型在多模态方面持续发展，当它们的“感受器”能感知更丰富的信息类型，AI科学家的能力将更加强大

4. 研究中所有实验都直接成功了

- 我们尚未测试AI是否能系统地解决科学实验中可能出现的问题并检测异常。这将是我们未来研究的方向