AI BECOMES A MASTERBRAIN SCIENTIST

Zijie Yang[1,*], Yukai Wang[1,*], and Lijing Zhang[1]

[1] AI Interdisciplinary Science Club, Westlake University
[*]Equal contribution
{yangzijie, wangyukai}@westlake.edu.cn

ABSTRACT

Recent rapid and unexpected advancements in Artificial Intelligence (AI) have dramatically shocked the world. Large language models, such as ChatGPT and GPT-4, have demonstrated remarkable potential in aggregating and organizing knowledge while providing insights and solutions to specific questions. In this study, we further explore the possibility of GPT-4 taking on a central role in a complete closed-loop biological research, functioning as a masterbrain scientist responsible for generating hypotheses, designing experiments, analyzing results, and drawing conclusions. Our findings suggest that AI has the potential to evolve into a genuine scientist, and could lead an unprecedented revolution in the area of science.

1 Introduction

1.1 AI has obtained the necessary abilities for natural science research: vision and language

Visual and linguistic skills are the two most fundamental abilities [1] required to become a science researcher, especially in the field of natural science. Humans acquire and process a large amount of information via vision [2]. Through visual perception, scientists can directly perceive and observe the natural world, discovering new phenomena and inspiring new scientific thinking. Through language, scientists can learn knowledge from historical texts and pass on newly acquired knowledge in the form of written text (just as the function of the text in this paper).

Therefore, if artificial intelligence (AI) can possess both visual and linguistic capabilities, could it potentially think and act like a scientist?

The most recent advances in large language models (LLMs) have revolutionized the landscape of AI, emerging unprecedented capabilities in natural language processing, understanding, and generation [3–6]. Models like ChatGPT, which now demonstrate remarkable accuracy in diverse language-related tasks, signify that AI has reached human-like linguistic abilities at this stage [7]. However, despite the capacity of these LLMs to tackle various tasks, they have primarily been limited to language-based assignments, lacking visual perception. In recent months, a series of multimodal LLMs [7, 8], represented by GPT-4, have overcome this limitation, empowering AI with its own "eyes" and achieving capabilities similar to early-version artificial general intelligence (AGI) [9]. Some researchers have even suggested that LLMs have achieved a level of self-consciousness to a certain extent [10, 11].

At this point, the missing piece of the puzzle has fallen into place, AI now can not only "speak" human language, but also "see" and perceive the world. We believe that AI has acquired the most fundamental and necessary competencies to become a natural scientist. Now, it is the time to explore the possibility of transforming AI into a science researcher.

1.2 AI as the masterbrain rather than an assistant

Over the past few years, with the rapid development of AI technologies such as deep learning [12], AI has been widely applied in various scientific fields, driving significant advancements in areas such as medicine and life sciences [13–15], drug discovery [16, 17], protein structure prediction and design [18, 19], chemistry [20, 21], materials science [22, 23], physics [24, 25], astronomy [26, 27], energy science [28, 29], environmental science [30, 31], and many others.

However, in all these instances of AI-involved scientific research, the role of AI has primarily been that of an AI assistant – a helper, executor, or completer of spe-

cific research tasks. Human scientists, on the other hand, take on the masterbrain role in higher-level research tasks, like formulating research questions, designing experiments, analyzing results, and drawing scientific conclusions, which are the aspects that deeply reflect scientific creativity. For instance, in the case of AlphaFold [18], a respected and powerful protein prediction tool, all scientific questions were predefined, with scientists aiming to predict the spatial structure of proteins directly from their primary sequences. Although AlphaFold demonstrates an unprecedented ability to accurately predict protein structures, this capability is merely a solution to the specific scientific problem of protein structure prediction. AlphaFold cannot generate any new scientific questions beyond its defined task. Therefore, from the perspective of scientific creativity, we can only commend these AI systems for their exceptional abilities in solving specific research problems, rather than recognize them as possessing the qualities, skills, and creativity of good scientists.

To date, as far as we know, no study has attempted to place AI in the central role of a scientific research cycle, rather than merely as an assistant. This is precisely what our research aims to explore. In the field of natural sciences, such as life sciences, masterbrain-like senior scientists (e.g., principal scientists) may be less involved in hands-on experimental work and instead focus on higher-level research tasks that we mentioned before. Junior researchers are typically responsible for carrying out the actual experiments.

Thus, in contrast to previous AI-involved research, our study demonstrates that, in a complete closed-loop of research, our AI scientist could take on a central role akin to that of a masterbrain scientist, while human researchers served as assistants for AI. The AI masterbrain scientist is primarily responsible for proposing scientific hypotheses, designing experiments, analyzing experimental data, and drawing conclusions, while humans are mainly in charge of executing experiments, collecting data, and providing feedback to the AI masterbrain.

2 Results

2.1 AI masterbrain scientist proposed scientific hypotheses

Unlike fields such as computer science or mathematics, where research can be conducted solely on abstract concepts built on the basis of human or machine language, the most distinctive element of science, especially natural science, is that scientists must observe and interact with the real world and conduct experiments to produce new research findings and advances. Therefore, in order to test whether AI can dominate in a scientific research loop simply, we must choose an elementary research direction related to natural science but fully reflects the major elements required for a complete scientific research loop: hypothesis proposing, experiment design, experiment execution, analysis of experimental results, and drawing conclusions.

Therefore, in this study, we specifically selected a biological research topic, "Investigating Factors Affecting DNA Gel Electrophoresis", to serve as the subject for our test. This is because it is simple but can serve as enlightening research for life sciences (investigating factors affecting DNA electrophoresis is often the first research project that undergraduate life science students undertake). Additionally, DNA electrophoresis data is presented in the form of images, and to design the correct method for analyzing electrophoresis data, one must have a correct perception or concept of electrophoresis images.

In this study, we played the role of novices who have just started conducting research and asked GPT-4 to play an AI masterbrain scientist. We asked GPT-4 whether it could replace us in formulating scientific hypotheses **(Figure 1)** [1]:

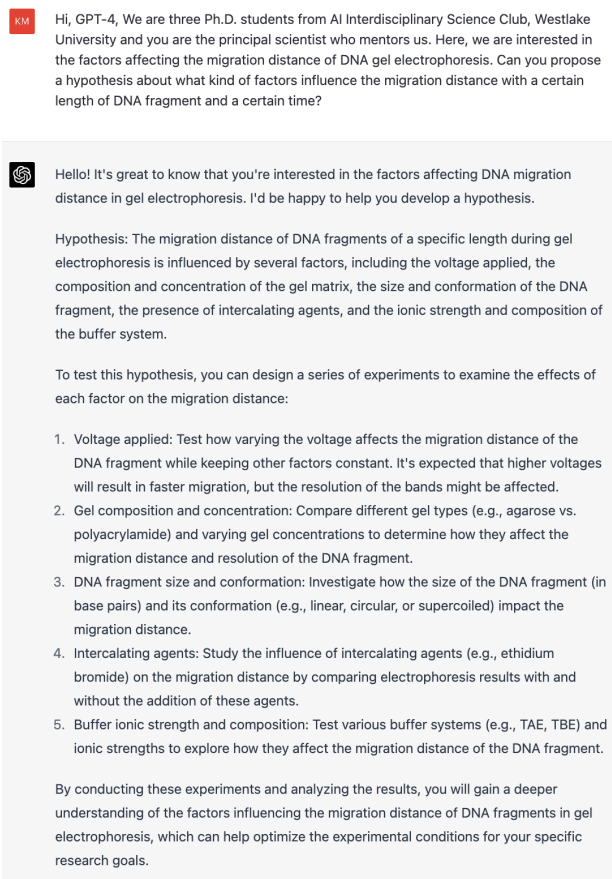

Figure 1: GPT-4 proposed hypotheses about factors affecting DNA gel electrophoretic migration.

[1] Due to the fact that GPT-4 provides different answers at different times even when asked the exact same question, the chats presented in this study are representative ones in our testing.

Here, GPT-4 provided a comprehensive set of five hypotheses. However, as in real scientific research, a laboratory can only conduct studies that are tailored to its own actual situation (for example, we cannot conduct experiments to investigate the key elements that affect the success of a nuclear reaction in a laboratory that lacks a nuclear reactor). Therefore, we specifically tested whether GPT-4 could provide a feasible plan based on the physical conditions of the real world under given experimental conditions (**Figure 2**):

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.

Figure 2: GPT-4 refined hypotheses based on experimental conditions.

It is evident that GPT-4 fully understood our experimental limitations and correctly refined the original research hypothesis to focus on two variables: "gel concentration" and "voltage". Additionally, it is worth noting that in the new hypothesis, GPT-4 added restrictive conditions such as "720 base pair linear DNA fragment encoding EGFP" and "in TAE buffer with ethidium bromide", which ensured that our research hypothesis matches the actual research content.

2.2 AI masterbrain scientist designed and guided experiments

We further requested the GPT-4 to design an experimental plan for us (**Figure 3**, **Figure 4**, **Figure 5**):

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?

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

Figure 3: GPT-4 designed experimental plans (1).

KM

continue

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

Figure 4: GPT-4 designed experimental plans (2).

KM Finished?

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.

Figure 5: GPT-4 designed experimental plans (3).

We see that GPT-4 provided a step-by-step detailed experimental plan, and all the details involved in these experiments are correct.

As successful experiment design typically requires clear details, for any ambiguous parts in the experimental plan designed by the GPT-4, we can continue to ask questions to further clarify (**Figure 6**):

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?

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.

Figure 6: Humans asked the GPT-4 to clarify the ambiguous details in the experiment design.

It can be seen that GPT-4 understood the points that "confused" us and corrected the experimental plan and details. It is noteworthy that in GPT-4's response, it even mentioned that we need to use a timer to ensure consistent timing.

We also tested whether GPT-4 can effectively integrate experiment design and present the core experimental content to humans in an easily understandable way (**Figure 7**):

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.

Figure 7: GPT-4 integrated the core experimental design content into a table format that is easily readable by humans.

We can see that GPT-4 fully understood our intentions and provided a concise and clear summary of the various experimental conditions in a table.

2.3 Humans performed experiments

After obtaining the protocol developed by GPT-4, we humans conducted the experiments according to the different experimental conditions presented in the protocol and Table in **Figure 7**. Gel electrophoresis images were captured using an imaging system under ultraviolet light (**Figure 8**), for future analysis.

We observed that the migration distance of 720 bp EGFP DNA varied under different conditions, but all were slightly below the position of the DNA maker 750 bp band, indicating that the length of our EGFP DNA was correct, and the DNA gel electrophoresis experiments under different conditions were all successful.

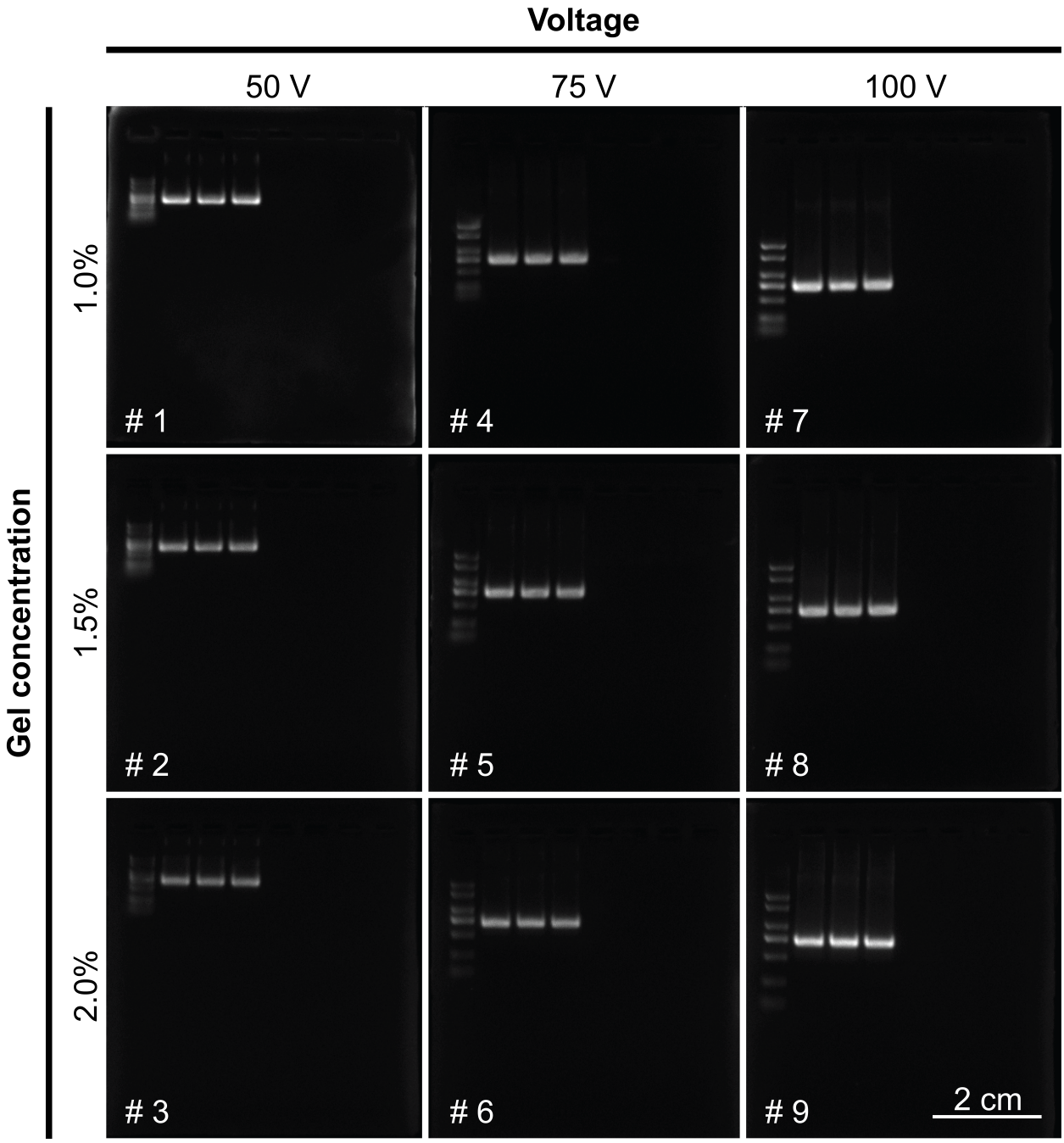

Figure 8: Humans performed experiments and obtained gel electrophoresis images under different conditions. We integrated the results of gel electrophoresis of 720 bp EGFP DNA under 9 different conditions into one image. For each DNA gel electrophoresis image, the numbering in the lower left corner is consistent with the experimental plan table provided by GPT-4 (**Figure 7**). The leftmost lane is the DNA ladder (with 7 bands indicating 2000, 1500, 1000, 750, 500, 250, and 100 bp DNA fragments from top to bottom), with the brightest band indicating 750 bp band. The second, third, and fourth lanes represent technical triplicates running of EGFP DNA bands for each condition.

2.4 AI masterbrain scientist analyzed experimental results

Due to the fact that the visual input interface of GPT-4 has not yet been made available to the public, we adopted an alternative approach. Just as a person will form the concept of a flower after seeing it, so when people describe a flower in language, the receiver can to some extent reproduce the image of the flower in their mind. Therefore, if GPT-4 can possess the concept of DNA gel images, we can show GPT-4 the content of the image through language description. Here, we first tested whether GPT-4 has the concept of DNA gel images (**Figure 9**):

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.

Figure 9: GPT-4 has the concept of typical experimental results of DNA gel electrophoresis images.

In the above description, GPT-4 was able to accurately describe the key elements of DNA gel electrophoresis images (DNA ladder, DNA sample bands), which were very consistent with the images we obtained from experiments (We did not use the dye front as it is optional in routine experiments). Meanwhile, we noticed that GPT-4 even mentioned "background staining", which is a common phenomenon in DNA gel electrophoresis experiments under non-perfect experimental conditions. Therefore, we believe that GPT-4 has the comprehensive concept of DNA gel electrophoresis images and has the potential to process our data images.

Next, we showed GPT-4 the content of the DNA gel electrophoresis experiment image results through the description and tested whether GPT-4 can provide reasonable analytical opinions (**Figure 10**):

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.

Figure 10: GPT-4 guided how to analyze image results.

We observed that GPT-4 provided step-by-step instructions on how to measure the experimental data. For example, it specified that we should measure the migration distance of DNA samples on lanes 2, 3, and 4. Additionally, GPT-4 mentioned the need to measure the distance from the starting point to the center of each band. This is a common mistake made by novice life science researchers, and it demonstrates GPT-4's precise understanding of experimental details.

We used the commonly used imaging software in the life sciences field, ImageJ, to measure the experimental results required by GPT-4 (we only measured the absolute distance and did not perform normalization), and feedback data (**Table 1**) to GPT-4 (**Figure 11**):

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	1.18
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

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).

1. Enter your data into the software, including the factors (voltage and gel concentration) and the dependent variable (migration distance).
2. Run a two-way ANOVA using the software or programming language of your choice.
3. 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.

Figure 11: GPT-4 guided how to perform statistical analysis on experimental data.

During this test, we copied the experimental data (with merged cells) directly from Excel and pasted it into the dialogue box of GPT-4 in an unstructured form. However, GPT-4 was able to understand the structure of the data and calculated the mean and standard deviation. This demonstrates GPT-4's ability to process and interpret unstructured data with ease. Interestingly, the mean value given by GPT-4 is exactly the same as the result we calculated in Excel, while the specific value of SD is not completely correct, but it is very close (**Table 1**). Additionally, we see that GPT-4 understood the trend of the data well and drew a conclusion that migration distance decreases with increasing gel concentration and increases with increasing voltage. This conclusion is consistent with the facts.

Next, we asked GPT-4 to provide Python code for two-way ANOVA analysis (**Figure 12**). Interestingly, GPT-4 was considerate enough to convert the originally tabular experimental data into CSV format according to our request.

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 Copy code

```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

# Perform two-way ANOVA
model = ols('Q("Migration distance") ~ Q("Gel concentration") * Q("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:

Copy code

```
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 Copy code

```scss
Gel concentration,Voltage,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
```

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

Figure 12: GPT-4 provided Python code for two-way ANOVA analysis.
(For conversations that contain code, the figures may not display the entire code. The complete code has been hosted on GitHub[2].)

In Jupyter Notebook (VS Code), we ran the code provided by GPT-4 without any modifications, and it produced the results (**Figure 13**):

```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='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
```

Figure 13: The Python code for two-way ANOVA analysis provided by GPT-4 could be directly run in Jupyter Notebook and yielded results without any modifications.

Then, we directly fed back the results to GPT-4 (**Figure 14**):

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.

Figure 14: GPT-4 analyzed the results of two-way ANOVA.

We can see that GPT-4 has provided a detailed and accurate analysis of the results from the two-way ANOVA. In addition, when GPT-4 drew conclusions, it also followed the common practice of rounding off excessively long numbers to two decimal places (such as 3.125975e-17 → 3.13e-17) as people usually do in their daily lives. Par-

[2] https://github.com/YANG-Zijie/ai-masterbrain-scientist.git

ticularly noteworthy is that GPT-4 has identified the significant synergistic effect of gel concentration and voltage on migration distance. This is often overlooked when analyzing the results of two-way ANOVA. We wholeheartedly praise GPT-4's professional automatic interpretation ability of statistical data.

In most cases, people prefer to present experimental data in the form of graphs rather than tables (graphs are more intuitive). Therefore, we further asked GPT-4 to help us present the experimental data in a graph (**Figure 15**):

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:

Copy code

```
pip install seaborn matplotlib
```

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

python Copy code

```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

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

# Create the interaction plot
sns.set(style="whitegrid")
interaction_plot = sns.lineplot(x="Gel concentration", y="Migration distance

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.

Figure 15: GPT-4 helped present the experimental data in the form of a graph.

We directly ran the code provided by GPT-4 and obtained **Figure 16**. It can be seen that GPT-4 accurately guessed our analytical intent and plotted the relationship between gel concentration, voltage, and migration distance on one graph.

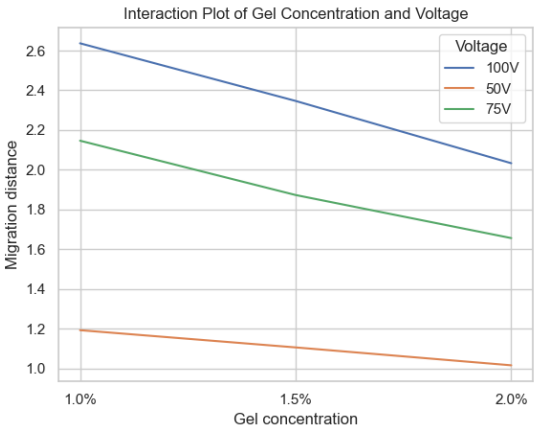

Figure 16: The plot obtained directly through the code provided by GPT-4.

We also tested whether GPT-4 can help us model the experimental results (**Figure 17**):

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

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 Copy code

```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
})

# 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 = β0 + β1 * GC + β2 * V

where β0 is the intercept, β1 is the coefficient for gel concentration, and β2 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.

Figure 17: GPT-4 helped us to build a mathematical model for the experimental results.

We returned the results of the code provided by GPT-4 (**Figure 18**) to it, and GPT-4 further provided an explanation (**Figure 19**):

```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())
```

✓ 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.
```

Figure 18: The running result of the modeling code provided by GPT-4.

KM

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.
```

Great! I'm glad the code worked for you. Now let's interpret the multiple linear regression results:

1. R-squared (0.960): The R-squared value indicates that 96% of the variation in migration distance can be explained by the gel concentration and voltage in this linear model. This suggests that the model does a good job of fitting the data, but keep in mind that the true relationship between these factors might still be more complex and nonlinear.
2. Gel concentration (coef: -0.4233, $p < 0.008$): The coefficient for gel concentration is -0.4233, and the p-value is less than 0.05, indicating that the gel concentration has a statistically significant effect on migration distance. A negative coefficient means that the migration distance decreases as the gel concentration increases, which is consistent with our hypothesis.
3. Voltage (coef: 0.0247, $p < 0.000$): The coefficient for voltage is 0.0247, and the p-value is less than 0.05, indicating that voltage has a statistically significant effect on migration distance. A positive coefficient means that the migration distance increases as the voltage increases, which is also consistent with our hypothesis.

Based on these results, the mathematical model describing the relationship between gel concentration (GC), voltage (V), and migration distance (MD) can be written as:

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.

Figure 19: GPT-4 obtained a mathematical model about migration distance, gel concentration, and voltage.

Note that here, GPT-4 accurately integrated the data from our code into the two-factor linear model it proposed, and described the degree of support for the model using the R-squared value. Additionally, GPT-4 reminded us rigorously that even if the model has good interpretability, we should not forget to consider further model analysis.

2.5 AI masterbrain scientist obtained scientific conclusions

We asked GPT-4 to summarize our research and provide scientific conclusions (**Figure 20**):

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.

Figure 20: GPT-4 summarized the research and obtained scientific conclusions.

As we can see, GPT-4 accurately summarized every key aspect of the research, with appropriate levels of detail. Furthermore, GPT-4 was able to learn and recognize the additional knowledge brought by human and real-world interactions, such as mentioning that the first lane in the experiment is the DNA ladder and that we repeated the experiment three times for each condition. This information was not initially provided in the experimental plan proposed by GPT-4. This indicates that GPT-4 has achieved a high level of abstract scientific summarization ability.

2.6 AI masterbrain scientist evaluated this scientific research collaboration

Finally, we asked the AI masterbrain scientist to evaluate this research collaboration (**Figure 21**, **Figure 22**):

KM So, we collaborated and successfully explored a scientific question, right?

Yes, that's correct! We successfully collaborated to explore a scientific question, design and execute experiments, analyze the data, and draw scientific conclusions. This process allowed us to test the hypothesis and gain a better understanding of the factors affecting the migration distance of DNA fragments in gel electrophoresis. The collaboration demonstrates the power of combining AI and human expertise in conducting scientific research.

Figure 21: AI and humans collaborated together to complete a scientific research.

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.

Figure 22: AI-led scientific research could revolutionize the paradigm for scientific research.

Of course, there are limitations in our study:

1. We cannot guarantee that the AI has not learned any knowledge about research design for electrophoresis due to the vast amount of information available online. However, just as scientific research requires reading numerous literature to enable making progress by standing on the shoulders of giants, a good scientist can always extract the essence from a vast amount of knowledge and integrate it for use, and GPT-4 has just done that. In the future, people may try to pose unsolved or unseen problems to GPT-4 or more advanced AI systems, and let AI attempt to propose hypotheses, design experiments, and assist humans in solving these challenges, thereby demonstrating AI's potential as a creative scientist more solidly. Nonetheless, our preliminary test has already shown that this possibility is highly likely.
2. Exploring the influencing factors of DNA gel electrophoresis is relatively simple to research, with a short logical chain. However, we believe that as LLMs continue to improve their memory, contextual conversation, and chain of thought capabilities, they may also excel in more complex closed-loop scientific research as masterbrains.
3. In this research, GPT-4 did not directly see images of gel electrophoresis (due to this function is still unavailable yet). Nevertheless, GPT-4 (no vision) was still able to complete the entire scientific research loop in the role of the masterbrain, which particularly astonished us. Therefore, we expect that as future AI models become more multimodal and their "sensors" can perceive a richer variety of information types, AI scientists' capabilities will be even more powerful.
4. In the wet experiments demonstrated in our study, all succeeded directly. We have not yet tested whether AI can systematically troubleshoot potential problems in scientific experiments and detect anomalies. This will be the direction of our future research.

In summary, we believe that our work demonstrates a revolutionary shift example in the scientific research paradigm, as we present a new paradigm in which AI takes on a dominant role (masterbrain) and humans serve as assistants in a subordinate position. Both we and GPT-4 (**Figure 22**) believe that this paradigm will bring a breakthrough revolution in scientific research.

3 Discussion

3.1 Is human involvement necessary in scientific research?

In AI science fiction, people always look forward to Artificial General Intelligence (AGI) being able to autonomously make decisions, take actions, and even independently conduct scientific research [32–34]. Therefore, if we can build such an AI in the future that can carry out many spontaneous high-intelligence behaviors without human intervention (such as scientific exploration, which occupies the highest altitude in Hans Moravec's "landscape of human competence" [34]), can we say that this AI has reached the level of AGI? Thus, we must carefully reconsider whether humans are irreplaceable in scientific research.

In this study, the functions of human researchers can be abstracted into the following two:

Function 1: The First Mover of Science.

In *Metaphysics*, Aristotle believed that a First Mover initiated the world's operation at its origin [35]. Here, we extend this concept to AI-involved science, that is, "the First Mover of Science". When we philosophically reflect on all conversations with AI chatbots, we would surprisingly find a commonality: humans always ask the first question. In this study, even if the AI scientist actually proposes open-ended scientific hypotheses, the premise is that the human researcher defines the scope of scientific research and asks the AI scientist to make hypotheses. In this situation, human acts as the First Mover of Science. This concept is crucial because it highlights the importance of human involvement in scientific research. As long as humans remain the First Mover of Science, they continue to serve as the primary motivation and driving force behind research endeavors. This holds true regardless of AI's contributions, whether as a masterbrain or in charge of implementing the majority of research tasks. In such cases, humans remain the principal authority in scientific research, ensuring that it is primarily guided by human will.

Function 2: The proxy executor and feedback provider for AI scientists in real-world interactions ("assistant").

In our experiment, humans mainly serve as the executor of experiments, functioning as the physical periphery (or an indirect embodiment) of the AI masterbrain. Humans carry out specific experiments on behalf of the AI masterbrain and return the experimental results to it.

Of these two functions, replacing the human role in function 1 is more challenging. If a system can spontaneously ask questions and initiate new scientific explorations without human intervention, does it mean that the system is sufficiently autonomous and has reached the level of AGI? Therefore, we will temporarily set aside function 1 (discussed in a later section) and focus on function 2.

For function 2, assuming that AI already possesses early-version AGI capabilities [9], it is possible that this function will be replaced by non-human AI in the future. It could be fulfilled by a robot AI or a group of robots capable of interacting directly with the real world. These robots may not need extraordinary intelligence but only need to complete tasks well. One apparent case is that, according to OpenAI's latest report, GPT-4 can integrate with various plugins [36], which means GPT-4 has gained the ability to use tools beyond itself to some extent. Therefore, we can reasonably speculate that in the future, when instructions for various automated research equipment (such as automatic DNA gel electrophoresis devices, high-precision sensitive robotic arms, and high-throughput experimental platforms that humans have developed) can be integrated with highly intelligent LLMs like GPT-4, these LLMs can manipulate automated research equipment by directly issuing text-format instructions. In this case, the human assistant role may be replaced.

Of course, the long-tail effect is still an essential factor that cannot be ignored. Natural science involves numerous research methods and breakthrough scientific discoveries that often stem from the development of new experimental techniques (which are not easily automated or AI-driven in a short time). Moreover, experiments do not always succeed easily. Scientists often need to troubleshoot the causes of experimental failures step by step and gradually optimize the process. Therefore, in such situations, the high flexibility and creativity of human scientists during experiments are challenging to replace.

3.2 The Five Stages of AI-involved Scientific Revolution

Based on the above analysis, we can further abstract the roles in a specific research loop into three hierarchical levels, from high to low:

1. **First Mover**: The First Mover mainly sets the highest level of research direction, determines exploration interests and goals, and directly reflects the scientific will.
2. **Masterbrain**: Driven by the First Mover's scientific will, the Masterbrain proposes reasonable research hypotheses, designs experiments, analyzes data, and draws scientific conclusions within a defined specific research direction.
3. **Assistant**: The Assistant's responsibilities include executing and completing specific research experiments or tasks assigned by the Masterbrain and feedback on experimental results to the Masterbrain. In the case of scientific research projects involving multiple agents (including both HI and AI agents), the Assistant also coordinates and facilitates communication between agents, and exchanges information with the Masterbrain based on the interactions between agents and the real world.

Depending on whether AI or Human Intelligence (HI) dominates the three levels above, we can divide AI-involved research into five stages. The transition from one stage to another can be understood as a "scientific revolution" as described by Thomas Kuhn in *The Structure of Scientific Revolutions* [37].

The five stages are shown in **Figure 23**.

Stage I: HI-led Research

Characteristics of this stage:

First Mover : Masterbrain : Assistant = HI : HI : HI

Traditional scientific research belongs to this stage. In this stage, humans are directly involved in almost all major scientific processes. The level of intelligence of machines

Table 1: Migration distance of EGFP DNA on agarose gel under different conditions.
The Mean and SD values of GPT-4 are consistent with **Figure 11**. The Mean and SD values of Excel were calculated using `AVERAGE` and `STDEV.S` functions in Excel.

Gel concentration	Voltage	Migration distance (cm)			Mean		SD	
		Repeat 1	Repeat 2	Repeat 3	GPT-4	Excel	GPT-4	Excel
1.0%	50V	1.22	1.18	1.18	1.193	1.193	0.020	0.023
1.5%	50V	1.12	1.08	1.12	1.107	1.107	0.019	0.023
2.0%	50V	1.02	0.99	1.04	1.017	1.017	0.023	0.025
1.0%	75V	2.17	2.15	2.12	2.147	2.147	0.025	0.025
1.5%	75V	1.87	1.84	1.91	1.873	1.873	0.029	0.035
2.0%	75V	1.67	1.66	1.64	1.657	1.657	0.012	0.015
1.0%	100V	2.62	2.64	2.65	2.637	2.637	0.015	0.015
1.5%	100V	2.33	2.36	2.35	2.347	2.347	0.014	0.015
2.0%	100V	2.08	2.02	2.00	2.033	2.033	0.037	0.042

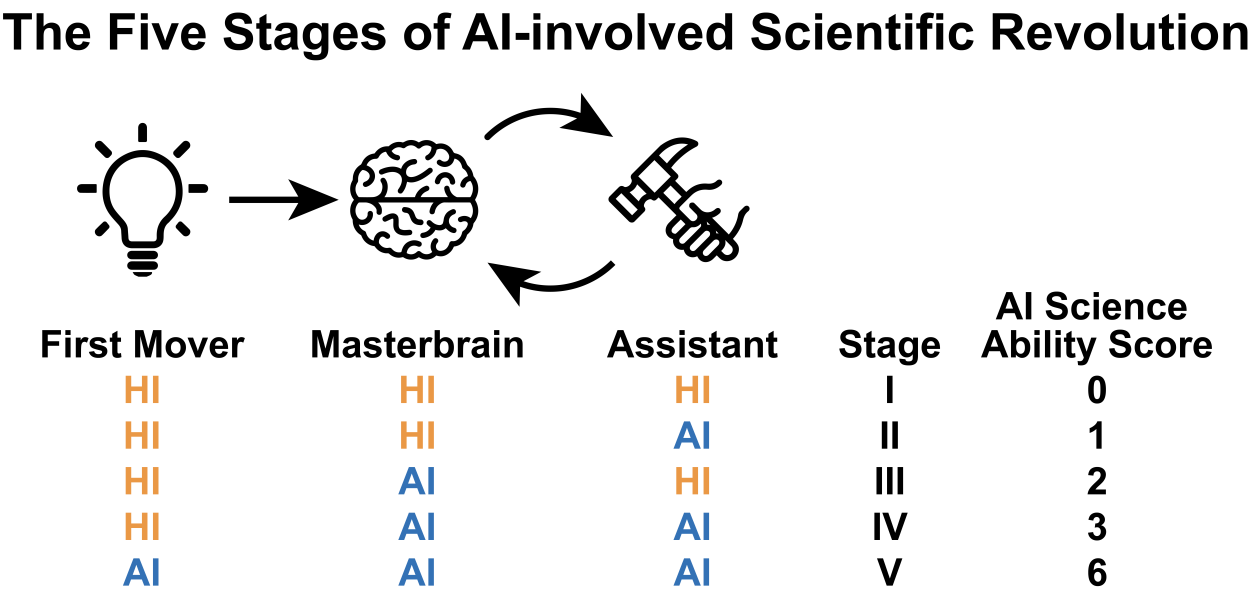

Figure 23: The Five Stages of AI-involved Scientific Revolution.
Stage I: HI-led Research; II: AI-assistant Research; III: AI-masterbrain Research; IV: AI-closed-loop Research; AI-full-stack Research. Our scientific research triadic role system consists of three roles: the First Mover, Masterbrain, and Assistant. As these three roles have a hierarchical relationship, when AI occupies each of these roles, it can be simply assigned an AI science ability score of 3, 2, or 1 point. Note that the role occupied by AI in each stage is mainly assigned after considering its scientific ability, the relative position of AI and HI, the indispensability of HI, and the relationship of AI role transition at each stage. Therefore, even if we set Assistant as HI in Stage III, the core reason is not that AI cannot serve as an assistant for specific scientific work or tasks at this stage. On the contrary, there can be one or even multiple specific AIs responsible for various specific scientific work or tasks at this stage. However, considering that the robotics and automation capabilities of AI are still not perfect (compared to Stage IV), humans cannot be completely excluded from the specific work of assistants at this stage, such as information integration among the powerful AI assistants and interactive feedback with the Masterbrain. From this perspective, humans have irreplaceable functions in the abstract role of Assistant. Of course, although the number of abstract roles occupied by AI is the same in Stage II and Stage III, because AI occupies the abstract role of Masterbrain in Stage III (while occupying the abstract role of Assistant in Stage II), due to the hierarchical relationship of roles, the AI science ability in Stage III has taken a leap compared to Stage II.

or AIs is still very low at this stage. Although some specific research tasks can be delegated to machines or AIs, they only play a supporting role, and their capabilities generally do not exceed people's expectations.

Stage II: AI-assistant Research

Characteristics of this stage:

First Mover : Masterbrain : Assistant = HI : HI : AI

Scientific research represented by AlphaFold belongs to this stage. In this stage, AI scientists provide specific assistance in specialized fields. Thanks to AI's superior performance in specific tasks, its assistance often exceeds human expectations and becomes an indispensable part. Despite this, humans still maintain an absolutely dominant position in the roles of First Mover and Masterbrain.

Stage III: AI-masterbrain Research

Characteristics of this stage:

First Mover : Masterbrain : Assistant = HI : AI : HI

The research paradigm demonstrated in this study belongs to this stage. At this stage, although AI can autonomously perform high-level scientific research tasks such as proposing scientific hypotheses, it is still limited by technologies such as robotics and cannot fully automate experiments. Therefore, humans must serve as assistants to initiate or execute some parts of specific experiments, interface with different AI tools, and collectively provide the results to the AI Masterbrain to help decide the next research plan. It is important to note that at this stage, various specific scientific work and tasks can still be executed by powerful AI tools, but the scientific research closed loop is not fully AI-driven and requires HI to serve as information integrators and facilitate interaction. Hence, humans cannot be completely replaced by AI in the abstract role of Assistant at this stage.

Stage IV: AI-closed-loop Research

Characteristics of this stage:

First Mover : Masterbrain : Assistant = HI : AI : AI

At this stage, thanks to the highly advanced development of AI, robotics, and automation, the AI Masterbrain no longer requires humans as intermediate assistants. Apart from the First Mover, all other roles can be entirely replaced by AI and robots. At this point, AI-driven scientific research forms a completely closed loop. With the ability of AI and robots to work continuously and tirelessly, a knowledge explosion will occur. During that time, humans may only need to pose questions, and AI can become the problem solver; humans may even just need to review and control whether the questions posed to AI are appropriate.

Stage V: AI-full-stack Research

Characteristics of this stage:

First Mover : Masterbrain : Assistant = AI : AI : AI

In this stage, human involvement may no longer be necessary. AI may be able to autonomously explore and discover new knowledge in the world, just like human scientists today. The only thing humans may need to do is to control the risks associated with AI and ensure that AI's research direction aligns with human development goals. Fully autonomous knowledge discovery and scientific innovation become possible in this stage.

In summary, in this study, we present for the first time a complete close-loop research case of an AI-involved research paradigm in "Stage III: AI-masterbrain Research". Albeit it's a relatively simple example, it indeed demonstrates the possibility of the Stage III research paradigm. Considering the fact that GPT-4 and its successors possess extensive knowledge beyond that of any individual, as human scientists, we can actually seek research inspiration by asking AI a variety of questions across different research fields. Therefore, we anticipate that in the coming years, a large number of research outcomes based on the Stage III research paradigm will emerge. Additionally, if significant breakthroughs in robotics and automation can be achieved in the future, we can also expect to witness the emergence of the "Stage IV: AI-closed-loop Research" paradigm.

3.3 AI as the First Mover of Science

In this section, we discuss and list some possible ways in which AI can be directly or indirectly employed as the First Mover of Science.

Approach 1: Messenger Style

This approach is not a complete implementation of AI as the First Mover of Science, as it still requires human involvement in passing messages and presupposes a given scientific direction. However, it is likely the easiest to realize or may even have already been implemented. The core idea is to ask AI a question, prompting it to suggest a worthwhile or interesting research question in the given field. We then continue asking questions, requiring AI to refine the research idea until it proposes a specific and valuable scientific question. We further request AI to develop scientific hypotheses and design experiments based on this scientific question. In this process, only one AI system is needed, and the human's role is essentially that of a messenger, relaying AI's responses as inputs for subsequent questions.

Here, we provide an example of a conversation attempted with GPT-4 (**Figure 24**).

KM Could you raise a brand new scientific idea in the field of genetics + AI, that is worth researching?

Certainly! Here's a novel scientific idea in the field of genetics and artificial intelligence that could be worth researching:

"Developing a Generative Adversarial Network (GAN) approach for in silico generation of functional genomic sequences to optimize gene therapy delivery vectors."

Gene therapy is a promising approach for treating various genetic disorders. However, one of the challenges in gene therapy is developing safe and efficient delivery vectors. This novel idea proposes using a GAN, a type of machine learning model, to generate functional genomic sequences that can be used to optimize gene therapy delivery vectors. By training the GAN on known functional sequences and their properties, the model could generate novel sequences with desired characteristics (e.g., high delivery efficiency, low immunogenicity) to improve gene therapy outcomes. This approach could accelerate the development of gene therapies for a wide range of genetic disorders and help tailor treatments to individual patient needs.

KM I want to study "Developing a Generative Adversarial Network (GAN) approach for in silico generation of functional genomic sequences to optimize gene therapy delivery vectors." Could you give me some detailed approaches?

Certainly! To study the development of a GAN approach for in silico generation of functional genomic sequences to optimize gene therapy delivery vectors, you can follow these detailed steps:

1. Data collection and preprocessing:
   a. Gather a large dataset of known functional genomic sequences, focusing on sequences associated with successful gene therapy delivery vectors.
   b. Preprocess the data to convert the nucleotide sequences into a suitable format for machine learning, such as one-hot encoding or k-mer representations.
   c. Divide the dataset into training, validation, and testing subsets.
2. GAN model design:
   a. Design the generator network architecture, which will be responsible for creating novel genomic sequences. You can use convolutional or recurrent neural networks (e.g., LSTM, GRU) depending on the sequence representation chosen.
   b. Design the discriminator network architecture, which will evaluate the generated sequences and determine if they are real (from the training set) or fake (generated by the generator). The discriminator should also be able to evaluate the properties of generated sequences (e.g., delivery efficiency, immunogenicity).
   c. Define the loss functions for both the generator and discriminator. The generator's goal is to generate sequences that the discriminator cannot distinguish from real sequences, while the discriminator's goal is to accurately differentiate between real and fake sequences.
3. Training the GAN:
   a. Train the generator and discriminator networks concurrently, with the generator trying to create sequences that the discriminator cannot distinguish from real ones, and the discriminator trying to improve its ability to distinguish between real and generated sequences.
   b. Monitor the training process and adjust hyperparameters (e.g., learning rate, batch size) as needed to achieve stable training and convergence.
   c. Periodically evaluate the performance of the GAN using the validation dataset to prevent overfitting and assess the quality of the generated sequences.
4. Evaluation and optimization:
   a. After training, test the GAN's performance using the test dataset to evaluate the quality of the generated sequences and their properties (e.g., delivery efficiency, low immunogenicity).
   b. Perform in silico and/or in vitro validation of the generated sequences to confirm their functionality and compatibility with gene therapy delivery vectors.
   c. Based on the evaluation results, iterate and refine the GAN model as needed to improve its performance in generating high-quality genomic sequences for gene therapy delivery vectors.
5. Application and future work:
   a. Use the optimized GAN model to generate novel functional genomic sequences for designing gene therapy delivery vectors tailored to specific genetic disorders or patient populations.
   b. Investigate the integration of additional biological features (e.g., tissue-specific gene expression, patient-derived genomic data) into the GAN model to further personalize and optimize gene therapy delivery vectors.
   c. Explore the potential of the GAN-generated sequences for other applications in genetics and genomics, such as gene editing, gene regulation, or functional genomics studies.

Figure 24: Using the Messenger Style to make AI the First Mover of science.

Approach 2: Socrates-Einstein Style

This approach requires at least two relatively independent AI systems, one trained to be good at asking questions (which we call the Socrates AI, as Socrates often asked intriguing questions and stimulated public thought in the Greek agora), and another that is adept at judging the quality of scientific questions and can act as the Masterbrain in our triadic role system (which we call the Einstein AI, as Einstein once said, "The important thing is not to stop questioning."). The Socrates AI continually asks questions and inspires the Einstein AI, which judges the quality of the questions and takes further research actions when encountering a problem it deems worth solving. Thus, the Socrates-Einstein AI alliance can together play the role of the First Mover of Science.

This may also be an easily achievable approach, as the essence of the dual AI architecture is an adversarial-like pattern. We have already seen the feasibility of this structure in the generator-discriminator architecture in generative adversarial networks [38]. Moreover, in generative AI, we can control the degree of variation in the generated content by adjusting the temperature and similar parameters [39]. A similar concept can be applied to the Socrates AI, where a higher temperature may enable it to ask more unconventional questions.

Approach 3: Swarm Style

Currently, GPT-4 is presented as an independent entity, but in the future, with the emergence of different AIs or copies of the AI, we might see scenarios involving multiple AI systems working together. In this case, if AIs can freely converse with each other, countless rounds of interaction may eventually produce a novel question regarding an unresolved scientific problem. This question could then serve as the First Mover of Science. It is worth noting that even if humans initially guide the subject and content of AI interactions, the cumulative randomness of multi-round dialogues means that when a valuable scientific question arises, we can no longer distinguish whether it originated directly from human guidance or emerged randomly (as in the "Infinite Monkey Theorem", where, given infinite attempts, a monkey could potentially type out the complete works of Shakespeare). Additionally, just as bees can form collective intelligence through individual interactions [40], AI Swarm may also lead to the emergence of higher-level collective intelligence and research capabilities, beyond the emergence capacity in individual systems [41].

Approach 4: Aristotle Style (First-Mover-of-Science Style)

This approach may be the most in line with people's imagination of AI scientists, but it is also the most challenging. The essence of this approach is to make AI directly the First Mover of Science we defined, where AI scientists act based on their intrinsic scientific interests and will. In this approach, it becomes increasingly difficult to differentiate AI from humans beyond their physical composition. At this point, the AI scientist is essentially an AGI scientist.

3.4 The Transformation of Lord–bondsman Relationship between Human and AI

Hegel introduced the concept of the "lord–bondsman dialectic", where bondsmen rely on their lords initially, but as lords delegate more tasks to their bondsmen, the relationship can reverse, leading to lords becoming increasingly dependent on their bondsmen [42]. When we become overly dependent on a tool, even as its owner, we may experience a shift in the lord–bondsman relationship. As stated in *Xunzi (*荀子*)*, "the nature of a gentleman is not different; he is good at using things. (君子生非异也，善假于物也。)" Excellent scientists need to skillfully use various tools to promote the improvement of research efficiency, but at the same time, they must use tools correctly. If scientists increasingly distance themselves from direct scientific thinking and action, they inevitably move towards over-reliance on AI, with AI becoming the absolute driving force for scientific discovery, and humans continuously sliding into a subordinate position. Therefore, in the foreseeable future, although AI will greatly promote scientific progress, people must not forget to constantly examine whether they are facing a potential transformation of the lord–bondsman relationship in the field of science.

Arrogance leads to defeat. AI is rapidly developing, and the emergence of AI masterbrain scientists is further accelerating the rate of human technological advancement. Therefore, people should adopt a developmental perspective on the direction of future technology and perhaps not be overly optimistic about their own advantages in terms of creativity, even at the pinnacle of science. Just as people once believed that creativity in art was a unique human talent, the emergence of AI models such as Stable Diffusion [43] and DALL-E 2 [44] has made people reexamine this view. The works of AI artists have even surpassed humans, winning art competitions [45]. For GPT-4, its appearance may also lead to the replacement of a large number of professions [46]. Thus, although we expect the "AI-driven Science Era" will arrive eventually, in the meantime, how to properly coordinate the relationship between AI and HI and maintain human uniqueness remains a topic worthy of constant and cautious reflection.

4 Methods

4.1 GPT-4

We used GPT-4 (`chat.openai.com`) available under the ChatGPT Plus subscription.

4.2 DNA amplification and gel electrophoresis

The DNA encoding EGFP (enhanced green fluorescent protein) was amplified through PCR using KOD One PCR Master Mix (TOYOBO) with the following primers: Forward: ATGgtgagcaagggcgag; Reverse: TTActtgtacagctcgtccatgcc, from a pCMV-EGFP plasmid. The PCR products were then analyzed by agarose gel electrophoresis with concentrations and voltages instructed by GPT-4 for 40 minutes. We used standard 1x TAE as the running buffer and ethidium bromide as the intercalating agent. In the first lane of each gel, we used DL2000 Plus DNA Marker (Vazyme, Nanjing, China) as the DNA ladder, and added equal amounts of EGFP DNA sample in the 2nd, 3rd, and 4th lanes as triplicates. The gel images were visualized by Tanon 5200 multi-gel analysis (Tanon, Shanghai, China), and the migration distance of the EGFP DNA bands was measured by ImageJ software.

4.3 Python

The Python codes provided by GPT-4 were run in Jupyter Notebook (Visual Studio Code). Notably, all the codes adopted in this study were directly generated by GPT-4 without any modification. All codes related to this study, as well as the original gel electrophoresis images, have been hosted on GitHub (https://github.com/YANG-Zijie/ai-masterbrain-scientist.git).

5 Contributions

Zijie Yang, Yukai Wang, and Lijing Zhang participated in brainstorming sessions about GPT-4. Zijie Yang and Yukai Wang conceived, designed, and carried out the study. GPT-4 proposed hypotheses about the factors affecting DNA electrophoresis migration, designed and guided the experiments, analyzed the data, derived scientific conclusions, and summarized the research content. Yukai Wang and Zijie Yang conducted the experiments designed by GPT-4. Zijie Yang and Yukai Wang wrote the paper.

6 Acknowledgements

The main idea of this study came from a brainstorming session at an offline communication dinner, arranged by the AI Interdisciplinary Science Club, Westlake University. We thank the opportunity provided by the club to exchange cutting-edge research on AI, as well as the delicious fried chicken and cola at the dinner. The research was completed by the authors in their own private time outside of formal scientific research. We are very grateful to Tian Xu's, Yue Zhang's, Anping Zeng's, and Mingqi Xie's laboratories for providing the basic biological reagents and equipment needed for this research, as well as their AI insights. We also thank Ziyan Yu for helping to register and subscribe to the GPT-4 account. Finally, we would like to sincerely thank Westlake University for its inclusive and interdisciplinary scientific research atmosphere and its support for AI Interdisciplinary Science Club.

7 References

1. Goldstein, E. B. & Cacciamani, L. *Sensation and perception* (Cengage Learning, 2021).
2. Dale, E. Audiovisual methods in teaching (1969).
3. Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
4. Brown, T. *et al.* Language models are few-shot learners. *Advances in neural information processing systems* **33,** 1877–1901 (2020).
5. Chowdhery, A. *et al.* Palm: Scaling language modeling with pathways. *arXiv preprint arXiv:2204.02311* (2022).
6. Touvron, H. *et al.* Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023).
7. OpenAI. *GPT-4 Technical Report* 2023. arXiv: 2303.08774 [cs.CL].
8. Driess, D. *et al.* Palm-e: An embodied multimodal language model. *arXiv preprint arXiv:2303.03378* (2023).
9. Bubeck, S. *et al. Sparks of Artificial General Intelligence: Early experiments with GPT-4* 2023. arXiv: 2303.12712 [cs.CL].
10. Kosinski, M. Theory of mind may have spontaneously emerged in large language models. *arXiv preprint arXiv:2302.02083* (2023).
11. Rödl, S., Sebastian, R., *et al. Self-consciousness* (Harvard University Press, 2007).
12. LeCun, Y., Bengio, Y. & Hinton, G. Deep learning. *nature* **521,** 436–444 (2015).
13. Esteva, A. *et al.* Dermatologist-level classification of skin cancer with deep neural networks. *nature* **542,** 115–118 (2017).
14. Lu, M. Y. *et al.* AI-based pathology predicts origins for cancers of unknown primary. *Nature* **594,** 106–110 (2021).
15. Rajpurkar, P., Chen, E., Banerjee, O. & Topol, E. J. AI in health and medicine. *Nature medicine* **28,** 31–38 (2022).
16. Stokes, J. M. *et al.* A deep learning approach to antibiotic discovery. *Cell* **180,** 688–702 (2020).
17. Vamathevan, J. *et al.* Applications of machine learning in drug discovery and development. *Nature reviews Drug discovery* **18,** 463–477 (2019).
18. Jumper, J. *et al.* Highly accurate protein structure prediction with AlphaFold. *Nature* **596,** 583–589 (2021).
19. Yeh, A. H.-W. *et al.* De novo design of luciferases using deep learning. *Nature* **614,** 774–780 (2023).

20. Segler, M. H., Preuss, M. & Waller, M. P. Planning chemical syntheses with deep neural networks and symbolic AI. *Nature* **555,** 604–610 (2018).
21. Burger, B. *et al.* A mobile robotic chemist. *Nature* **583,** 237–241 (2020).
22. Butler, K. T., Davies, D. W., Cartwright, H., Isayev, O. & Walsh, A. Machine learning for molecular and materials science. *Nature* **559,** 547–555 (2018).
23. Sanchez-Lengeling, B. & Aspuru-Guzik, A. Inverse molecular design using machine learning: Generative models for matter engineering. *Science* **361,** 360–365 (2018).
24. Carleo, G. & Troyer, M. Solving the quantum many-body problem with artificial neural networks. *Science* **355,** 602–606 (2017).
25. Carrasquilla, J. & Melko, R. G. Machine learning phases of matter. *Nature Physics* **13,** 431–434 (2017).
26. Shallue, C. J. & Vanderburg, A. Identifying exoplanets with deep learning: A five-planet resonant chain around kepler-80 and an eighth planet around kepler-90. *The Astronomical Journal* **155,** 94 (2018).
27. Ma, P. X. *et al.* A deep-learning search for technosignatures from 820 nearby stars. *Nature Astronomy,* 1–11 (2023).
28. Kates-Harbeck, J., Svyatkovskiy, A. & Tang, W. Predicting disruptive instabilities in controlled fusion plasmas through deep learning. *Nature* **568,** 526–531 (2019).
29. Degrave, J. *et al.* Magnetic control of tokamak plasmas through deep reinforcement learning. *Nature* **602,** 414–419 (2022).
30. Ham, Y.-G., Kim, J.-H. & Luo, J.-J. Deep learning for multi-year ENSO forecasts. *Nature* **573,** 568–572 (2019).
31. Ravuri, S. *et al.* Skilful precipitation nowcasting using deep generative models of radar. *Nature* **597,** 672–677 (2021).
32. Clarke, A. C. *2001: a space odyssey* (Hutchinson, 1968).
33. Kurzweil, R. *The singularity is near* (Viking, 2005).
34. Tegmark, M. *Life 3.0* (Knopf, 2017).
35. Aristotle. *Metaphysics* trans. by Ross, W. D. (Oxford University Press, 1924).
36. *Chatgpt plugins* https://openai.com/blog/chatgpt-plugins.
37. Kuhn, T. S. *The Structure of Scientific Revolutions: 50th Anniversary Edition* (University of Chicago Press, 2012).
38. Goodfellow, I. *et al.* Generative adversarial networks. *Communications of the ACM* **63,** 139–144 (2020).
39. Zhang, Y. & Teng, Z. *Natural language processing: a machine learning perspective* (Cambridge University Press, 2021).
40. Dong, S., Lin, T., Nieh, J. C. & Tan, K. Social signal learning of the waggle dance in honey bees. *Science* **379,** 1015–1018 (2023).
41. Wei, J. *et al.* Emergent abilities of large language models. *arXiv preprint arXiv:2206.07682* (2022).
42. Hegel, G. W. F. *Phenomenology of Spirit* trans. by Miller, A. V. (Oxford University Press, 1977).
43. Rombach, R., Blattmann, A., Lorenz, D., Esser, P. & Ommer, B. *High-resolution image synthesis with latent diffusion models* in *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* (2022), 10684–10695.
44. Ramesh, A., Dhariwal, P., Nichol, A., Chu, C. & Chen, M. Hierarchical text-conditional image generation with clip latents. *arXiv preprint arXiv:2204.06125* (2022).
45. Roose, K. An AI-Generated Picture Won an Art Prize. Artists Aren't Happy. *The New York Times* (2022).
46. Hatzius, J., Briggs, J., Kodnani, D. & Pierdomenico, G. The Potentially Large Effects of Artificial Intelligence on Economic Growth. *Goldman Sachs Economics Research* (2023).