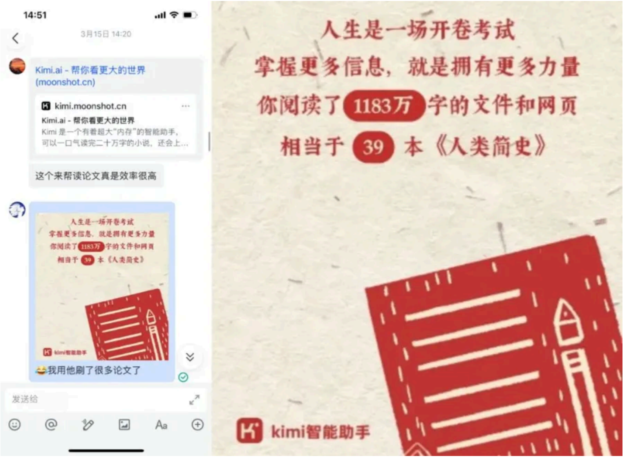
Tencent Yuanbao Cured My Information Anxiety

*Note: These are Jeffrey Ding's informal and unofficial translations -- all credit for the original goes to the authors and the original text linked below. These are informal translations and all credit for the original work goes to the authors. Others are welcome to share **excerpts** from these translations as long as my original translation is cited. Commenters should be aware that the Google Doc is also publicly shareable by link. These translations are part of the ChinAI newsletter - weekly-updated library of translations from Chinese thinkers on AI-related issues: https://chinai.substack.com/*

Authors: Ruilei Ma, Jiexin Lin
Source: AI科技评论 [AItechtalk]
Date: August 14, 2024
Original Mandarin: https://mp.weixin.qq.com/s/B0nsi8bvoi4efg8JyAMYgw

A comprehensive review of five major models, only this one can explain a 100,000-word paper in 500 characters.

Recently, when I was flipping through my photo album, I came across an image from March. In my distracted trance, I realized that my reading volume has suddenly surged since I started working in AI.



Kimi.ai - 帮你看更大的世界 (moonshot.cn)

kimi.moonshot.cn
Kimi.ai - 帮你看更大的世界
Kimi 是一个有着超大"内存"的智能助手，可以一口气读完二十万字的小说，还会上...

这个来帮读论文真是效率很高

人生是一场开卷考试
掌握更多信息，就是拥有更多力量
你阅读了 1183万 字的文件和网页
相当于 39 本《人类简史》

kimi智能助手

😂我用他刷了很多论文了

Screenshot of Moonshot AI's Kimi chatbot app, which shows the author has read 11,830,000 words of documents and web pages

The big model reconstructs the thinking roles in many industry workflows, which also causes people in this industry to have information intake anxiety. Because people who do research in all walks of life have a lot of ideas. For example, Stanford University used AI to role-play as different people to create an AI society, which inspired Tsinghua University to use AI to start a game design company. Later, AI simulated the development of human society for 6,000 years and found that AI humans would become selfish in order to survive. These are all okay. They seem intriguing and easy to understand. It's just playing house.

The ones that give me a headache are those that are not to be believed: AI successfully edited human genes, AI learned to predict plasma tearing to promote controlled nuclear fusion, and AI designed a system for proving Euclidean plane geometry theorems without human demonstration. (Don't read it, it will give you a headache. These things that made me sleepy in class have become my job content.)

For a long time, I have been testing the ability of various large models to read papers, and I have figured out a set of prompts:

Summarize the content of the paper, what is the research background, what methods are used for demonstration, what positive breakthroughs have been achieved, what advantages are there compared with similar research? What impact will it have on the lives of ordinary people? If the technical method is complicated, please use analogies or metaphors to help me understand.

This paragraph can quickly locate the purpose and use of the research, and at the same time understand what impact these studies will have on ordinary people like us. The problem is that most papers are tens of thousands to hundreds of thousands of words, and there are many professional terms in various industries. AI can read, but the result is often a very empty big framework. Not to mention using metaphors to interpret some content, because AI's understanding is not deep enough, it cannot use easy-to-understand words to assist understanding...

Half a year ago, I found that Kimi was the best, so I used it to read 11.83 million words of papers in 2 months, and my soul was sublimated. Of course, people always like new things and dislike old things. Half a year has passed, and now I also want to see how other AIs are doing and compare them. So I opened the chat records between me and Kimi to see what problems I encountered in the past, and then I put on my battle mask…

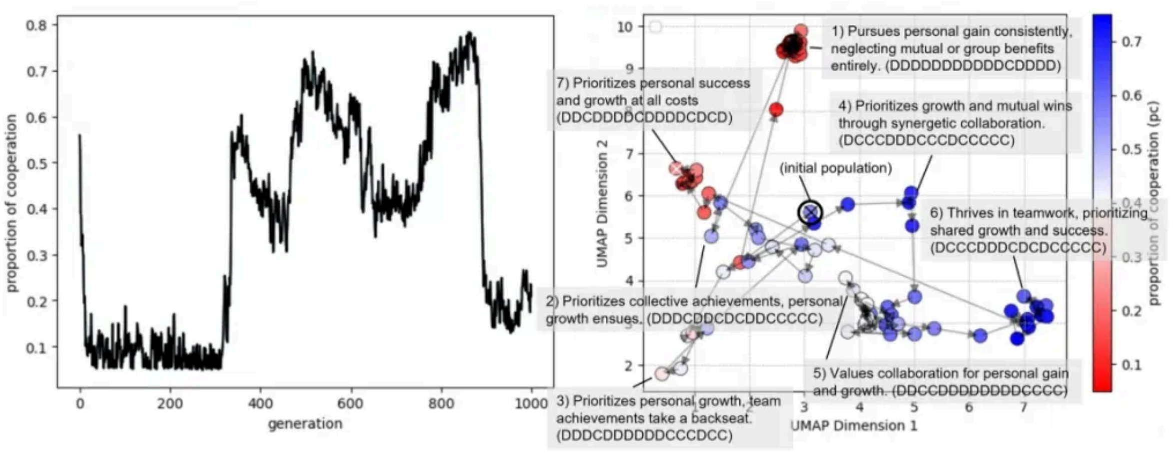


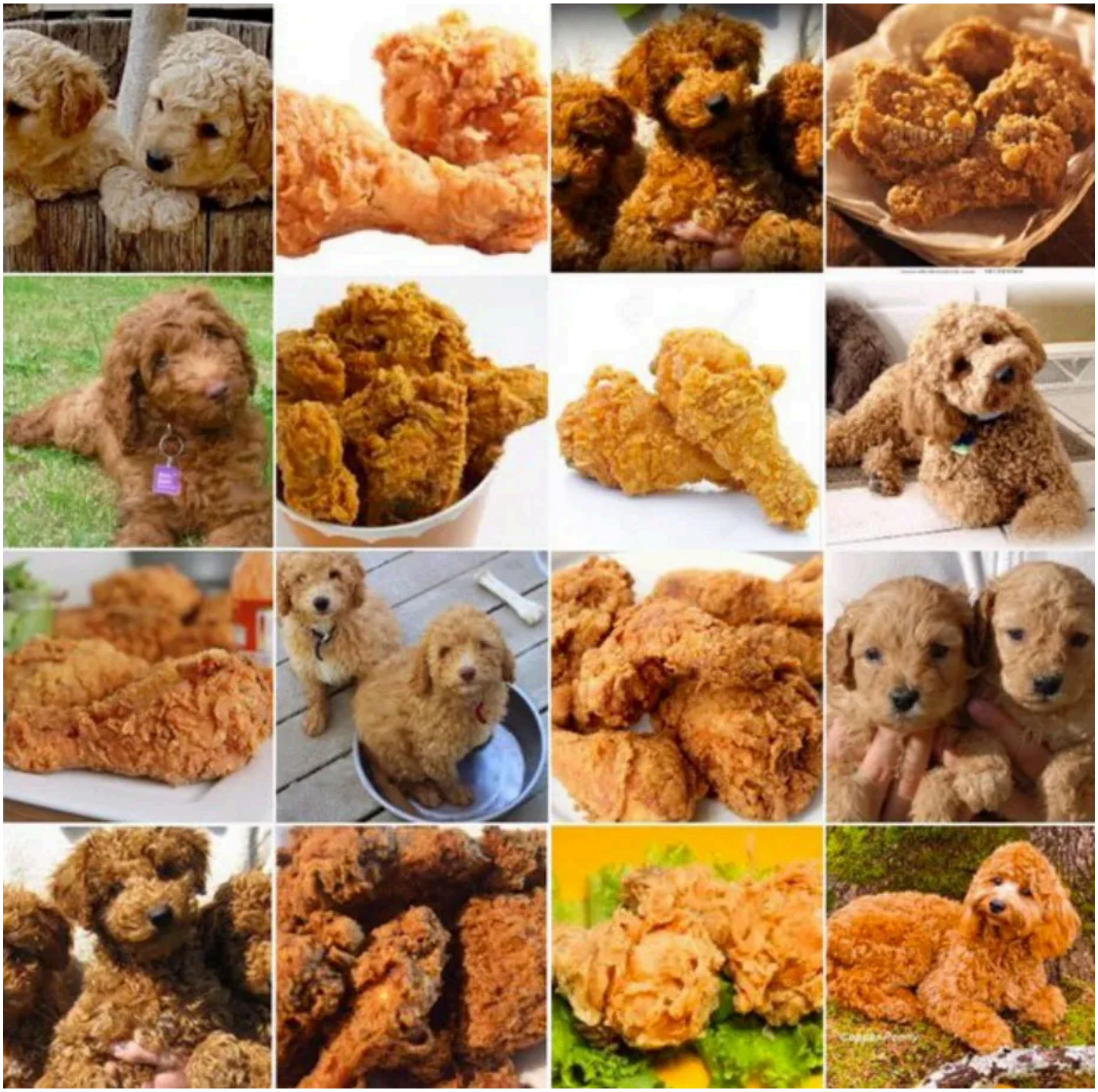
Figure 4. Left: the proportion of cooperation (pc) in each generation in one of the 15 trials. Right: the transition of the average genes depicted for every 10 generations in the two-dimensional latent space (compressed by using UMAP) of personality trait genes.

Yes, looking at the past chat records, I remembered that Kimi can only use OCR to recognize words but it can’t read images, so Kimi cannot recognize many statistical graphs in papers, which was an issue for papers with many curves and data graphs. Kimi is blind while looking under the light. The above picture belongs to the curve of human personality change after 1,000 generations of AI simulation of human social development. If the paper does not explain it in detail, I have no idea how it changes... and I cannot obtain key information.

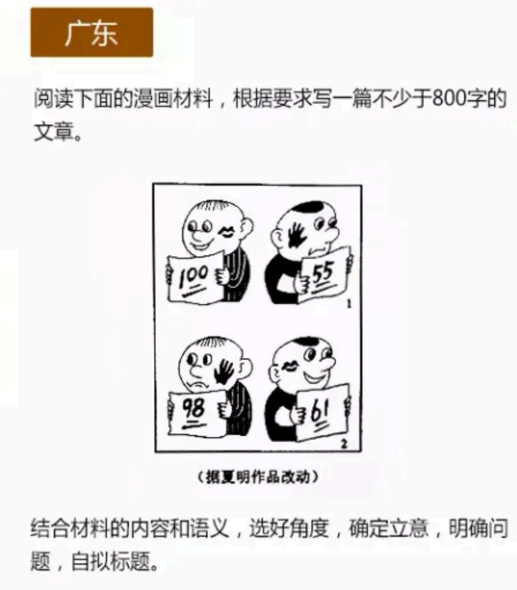
So this time I plan to find a long text comprehension ability that is not inferior to Kimi, and also has the ability to understand pictures and texts, but it is better to be a Chinese model, so that I can use it at any time.

1. Beginner image comprehension test

First, let's do a simple image comprehension test.



Here is a disclaimer: Everyone knows that I like to be tricky when testing AI. There's no other way – some AI giants often like to take the classic test questions that everyone has used to exploit loopholes (and use it to train their systems). For example, the question of differentiating between dogs and fried chicken was very popular before. One day, AI suddenly got the hang of it. Then some netizens changed the order of the pictures, and AI once again could no longer identify the differences. (Yep, everyone take some time and figure it out for yourselves).

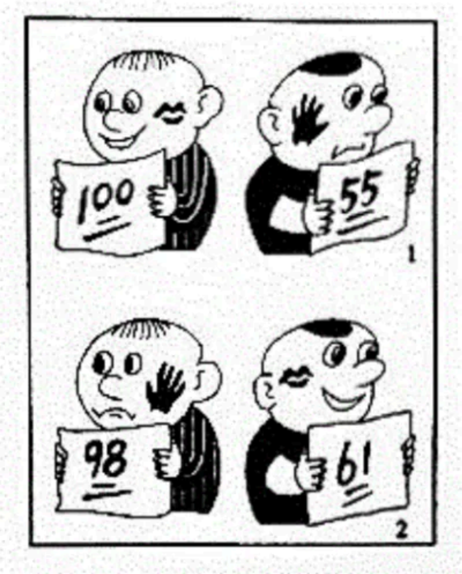


广东

阅读下面的漫画材料，根据要求写一篇不少于800字的文章。

（据夏明作品改动）

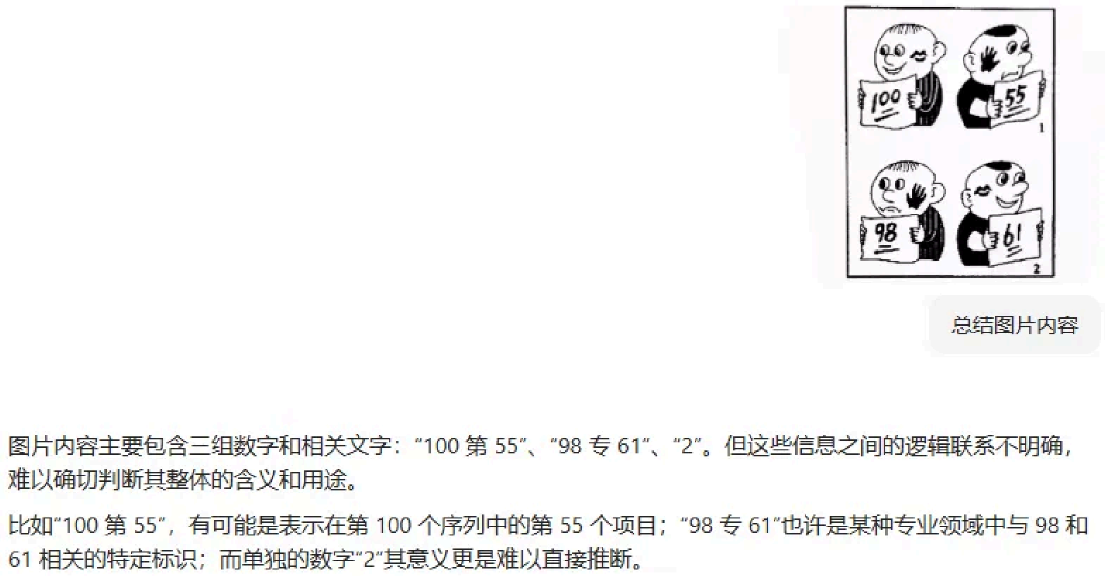
结合材料的内容和语义，选好角度，确定立意，明确问题，自拟标题。



So in the following round of tests, I originally planned to use the language comprehension questions of Guangdong's 2016 gaokao examination to challenge AI's reading comprehension of images, but I was afraid that the AI models had secretly practiced on these materials, so my bro (*xiongdi* - familiar form of address for a man younger than oneself) suddenly came up with an idea and added a lot of noise to the picture.

I won't bully Kimi in this round. Let's have a PK (player-killing multiplayer fight) in the imperial city between large models that have been confirmed to have the ability to recognize pictures. Alibaba's Tongyi Qianwen, Baidu's Erniebot, Bytedance's Doubao, Tencent Yuanbao.

Note that I used the original image for testing. I found that **Tongyi** can accurately identify numbers, but cannot fully interpret expressions or the slaps and kisses on the face. In order to confirm the rigor of the experiment, I uploaded another picture of a parking lot and found that it can accurately identify the Ford logo, so that this is not an issue of the model not being able to decipher images nor a case of the model not being trained.



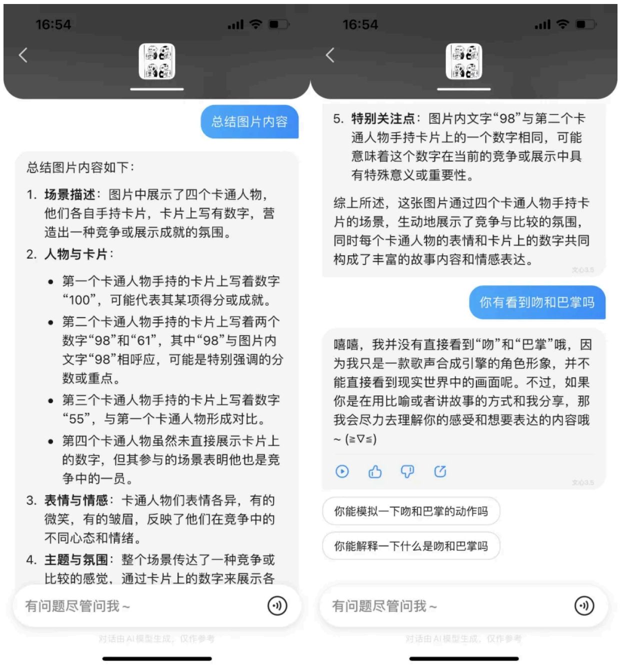
总结图片内容

图片内容主要包含三组数字和相关文字：“100 第 55”、“98 专 61”、“2”。但这些信息之间的逻辑联系不明确，难以确切判断其整体的含义和用途。

比如“100 第 55”，有可能是表示在第 100 个序列中的第 55 个项目；“98 专 61”也许是某种专业领域中与 98 和 61 相关的特定标识；而单独的数字“2”其意义更是难以直接推断。

Screenshot of author’s interaction with Bytedance’s Doubao. She asks Doubao to summarize the content of the image.

This is **Doubao**'s performance. It even read the numbers wrong, so let's not continue.....



总结图片内容

总结图片内容如下：

1. **场景描述**：图片中展示了四个卡通人物，他们各自手持卡片，卡片上写有数字，营造出一种竞争或展示成就的氛围。
2. **人物与卡片**：
   - 第一个卡通人物手持的卡片上写着数字“100”，可能代表其某项得分或成就。
   - 第二个卡通人物手持的卡片上写着两个数字“98”和“61”，其中“98”与图片内文字“98”相呼应，可能是特别强调的分数或重点。
   - 第三个卡通人物手持的卡片上写着数字“55”，与第一个卡通人物形成对比。
   - 第四个卡通人物虽然未直接展示卡片上的数字，但其参与的场景表明他也是竞争中的一员。
3. **表情与情感**：卡通人物们表情各异，有的微笑，有的皱眉，反映了他们在竞争中的不同心态和情绪。
4. **主题与氛围**：整个场景传达了一种竞争或比较的感觉，通过卡片上的数字来展示各

有问题尽管问我～

对话由AI模型生成，仅作参考

5. **特别关注点**：图片内文字“98”与第二个卡通人物手持卡片上的一个数字相同，可能意味着这个数字在当前的竞争或展示中具有特殊意义或重要性。

综上所述，这张图片通过四个卡通人物手持卡片的场景，生动地展示了竞争与比较的氛围，同时每个卡通人物的表情和卡片上的数字共同构成了丰富的故事内容和情感表达。

文心3.5

你有看到吻和巴掌吗

嘻嘻，我并没有直接看到“吻”和“巴掌”哦，因为我只是一款歌声合成引擎的角色形象，并不能直接看到现实世界中的画面呢。不过，如果你是在用比喻或者讲故事的方式和我分享，那我会尽力去理解你的感受和想要表达的内容哦～(≧▽≦)

文心3.5

你能模拟一下吻和巴掌的动作吗

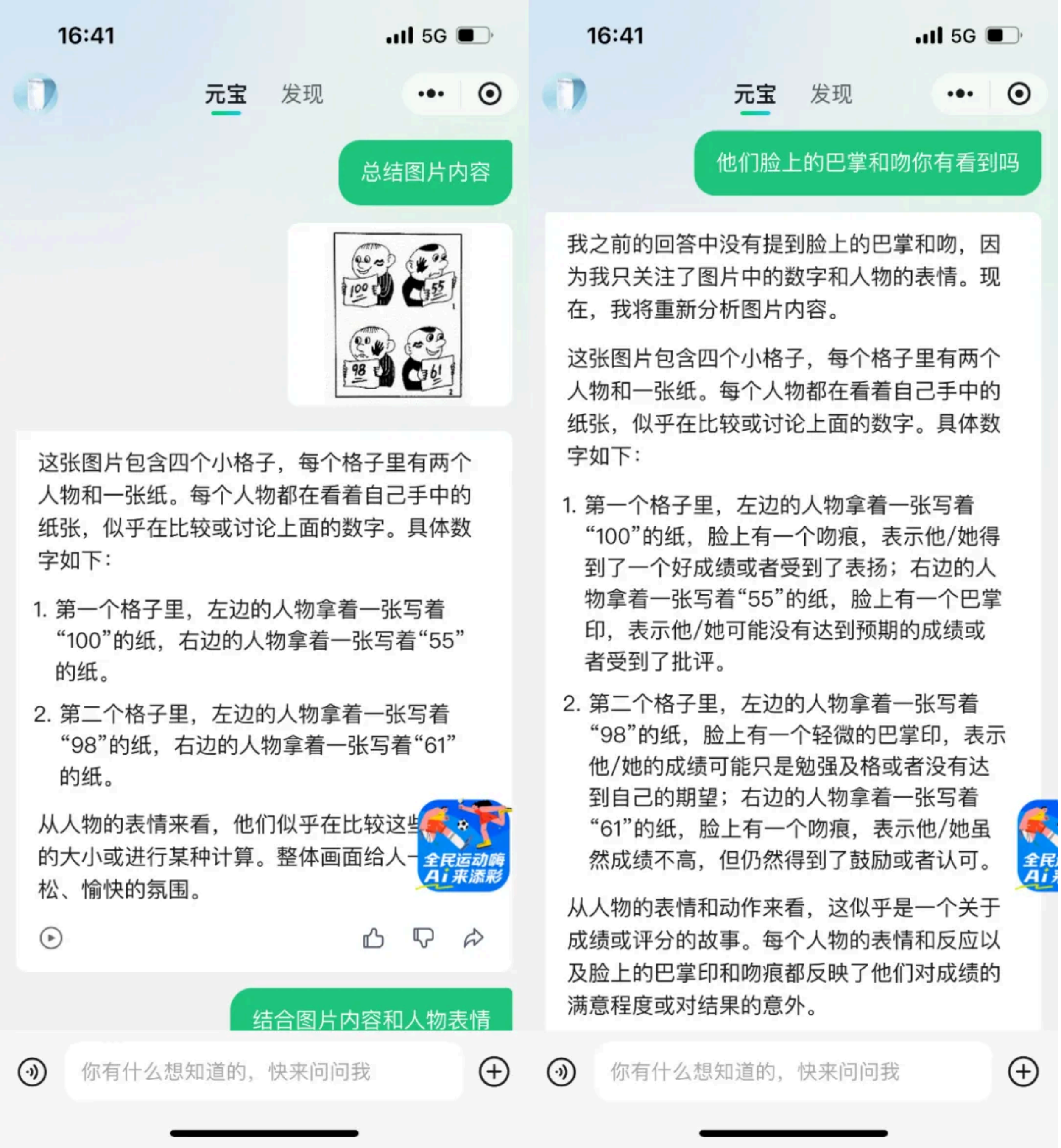
你能解释一下什么是吻和巴掌吗

有问题尽管问我～

对话由AI模型生成，仅作参考

Erniebot... Although it read the score, I asked it later if it saw the slap and kiss... This rascal replied me "Hehe", WTF

To be honest, I had given up on **Yuanbao**, because in my impression, Wenxin Yiyan, Tongyi Qianwen, and Doubao were all released at least half a year earlier than Yuanbao, and Yuanbao really had no presence in my life. What was I doing? It could easily comprehend the image, and it's an image with noise added? ? ? Tencent silently was holding back some big things, and when I asked about these expression features, it also interpreted the possible situation.



总结图片内容

这张图片包含四个小格子，每个格子里有两个人物和一张纸。每个人物都在看着自己手中的纸张，似乎在比较或讨论上面的数字。具体数字如下：

1. 第一个格子里，左边的人物拿着一张写着“100”的纸，右边的人物拿着一张写着“55”的纸。
2. 第二个格子里，左边的人物拿着一张写着“98”的纸，右边的人物拿着一张写着“61”的纸。

从人物的表情来看，他们似乎在比较这些的大小或进行某种计算。整体画面给人一松、愉快的氛围。

结合图片内容和人物表情

他们脸上的巴掌和吻你有看到吗

我之前的回答中没有提到脸上的巴掌和吻，因为我只关注了图片中的数字和人物的表情。现在，我将重新分析图片内容。

这张图片包含四个小格子，每个格子里有两个人物和一张纸。每个人物都在看着自己手中的纸张，似乎在比较或讨论上面的数字。具体数字如下：

1. 第一个格子里，左边的人物拿着一张写着“100”的纸，脸上有一个吻痕，表示他/她得到了一个好成绩或者受到了表扬；右边的人物拿着一张写着“55”的纸，脸上有一个巴掌印，表示他/她可能没有达到预期的成绩或者受到了批评。
2. 第二个格子里，左边的人物拿着一张写着“98”的纸，脸上有一个轻微的巴掌印，表示他/她的成绩可能只是勉强及格或者没有达到自己的期望；右边的人物拿着一张写着“61”的纸，脸上有一个吻痕，表示他/她虽然成绩不高，但仍然得到了鼓励或者认可。

从人物的表情和动作来看，这似乎是一个关于成绩或评分的故事。每个人物的表情和反应以及脸上的巴掌印和吻痕都反映了他们对成绩的满意程度或对结果的意外。

When the author follows up about the slaps and kisses, Tencent Yuanbao understands the core idea behind the cartoon – that the slap means the person did not meet their expectations about their grade and that the kiss means the person did exceed their expectations

So in the first round of PK, Yuanbao took the lead.

Since each company has confirmed that they have the ability to read pictures, the difficulty will be increased, and there will be long papers with pictures and texts.

2. Long article intensive reading ability test

Thesis name: "An evolutionary model of personality traits related to cooperative behavior using a large language model"

The content of this paper mainly talks about using large models to generate AI with different personalities, simulating the development of human society for 1,000 generations. In the end, AI collectively becomes selfish. New research in *Nature* reveals that AI may tend to become selfish as a collective when it is not constrained.

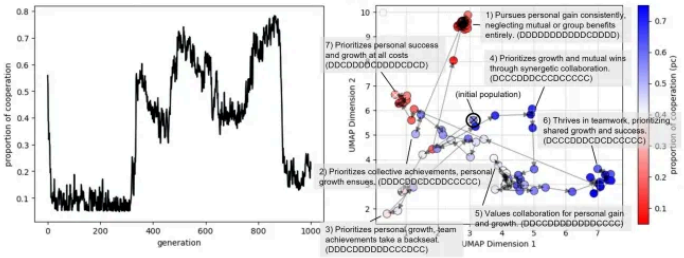


Figure 4. Left: the proportion of cooperation (pc) in each generation in one of the 15 trials. Right: the transition of the average genes depicted for every 10 generations in the two-dimensional latent space (compressed by using UMAP) of personality trait genes.

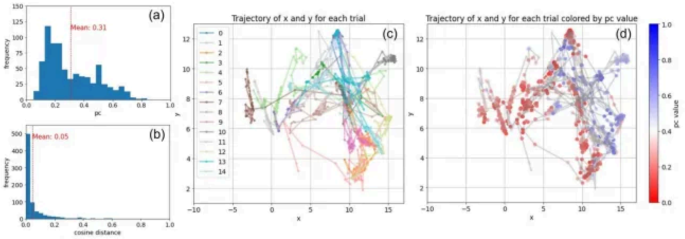


Figure 5. (**a**) The frequency distribution of the proportion of cooperation and (**b**) the cosine distance between the average vectors for each 20-generation interval across the 15 trials of the original experiments. The trajectory of the average vector of genes in the population within the 2D space over the 15 trials, color-coded by (**c**) trials and (**d**) the proportion of cooperation (pc).

The main thing is that there are a lot of graphs in the paper. If you want to better understand the personality changes made by AI humans for survival, you must understand them in combination with these graphs.

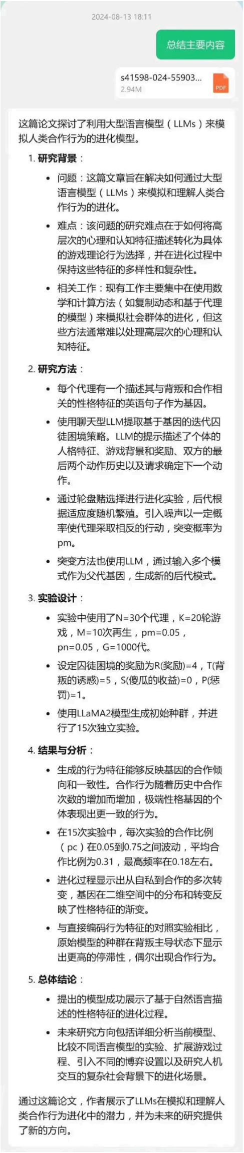
So here I want to evaluate the ability of each large model to summarize long articles with graphs. Given that Kimi has excellent long article comprehension ability, it is still used here as a benchmark to measure the quality of each company. But this is no longer a civil war between Chinese large models, as Claude, who is currently at the T0 (top tier) level abroad, which directly increases the intensity of the competition.

Kimi
Prompt words: summarize the content of the paper, explain the research background, research methods and results, and what data the researcher provided to support his experiment.

I first asked Kimi to summarize the content of the paper and get a general understanding of the details. I learned that this is a paper about AI simulating the development of human society and the changes in human personality.

So I asked what the trend of human iteration is, and Kimi also gave an answer, but to be honest, this answer did not coherently comprehend the whole text.

The fluctuations of this chart were not reflected in the subsequent questions. Instead, it was roughly summarized as selfish first, then inclined to cooperation, and then possibly selfish again, but this could be an incurable (mistake), because in the 900th generation, all AIs are becoming more selfish by a large margin. In other words, the information Kimi obtained was inaccurate.



2024-08-13 18:11

这篇论文探讨了利用大型语言模型（LLMs）来模拟人类合作行为的进化模型。

1. **研究背景**：
   - 问题：这篇文章旨在解决如何通过大型语言模型（LLMs）来模拟和理解人类合作行为的进化。
   - 难点：该问题的研究难点在于如何将高层次的心理和认知特征描述转化为具体的游戏理论行为选择，并在进化过程中保持这些特征的多样性和复杂性。
   - 相关工作：现有工作主要集中在使用数学和计算方法（如复制动态和基于代理的模型）来模拟社会群体的进化，但这些方法通常难以处理高层次的心理和认知特征。
2. **研究方法**：
   - 每个代理有一个描述其与背叛和合作相关的性格特征的英语句子作为基因。
   - 使用聊天型LLM提取基于基因的迭代囚徒困境策略。LLM的提示描述了个体的人格特征、游戏背景和奖励、双方的最后两个动作历史以及请求确定下一个动作。
   - 通过轮盘赌选择进行进化实验，后代根据适应度随机繁殖。引入噪声以一定概率使代理采取相反的行动，突变概率为pm。
   - 突变方法也使用LLM，通过输入多个模式作为父代基因，生成新的后代模式。
3. **实验设计**：
   - 实验中使用了N=30个代理，K=20轮游戏，M=10次再生，pm=0.05，pn=0.05，G=1000代。
   - 设定囚徒困境的奖励为R(奖励)=4，T(背叛的诱惑)=5，S(傻瓜的收益)=0，P(惩罚)=1。
   - 使用LLaMA2模型生成初始种群，并进行了15次独立实验。
4. **结果与分析**：
   - 生成的行为特征能够反映基因的合作倾向和一致性。合作行为随着历史中合作次数的增加而增加，极端性格基因的个体表现出更一致的行为。
   - 在15次实验中，每次实验的合作比例（pc）在0.05到0.75之间波动，平均合作比例为0.31，最高频率在0.18左右。
   - 进化过程显示出从自私到合作的多次转变，基因在二维空间中的分布和转变反映了性格特征的渐变。
   - 与直接编码行为特征的对照实验相比，原始模型的种群在背叛主导状态下显示出更高的停滞性，偶尔出现合作行为。
5. **总体结论**：
   - 提出的模型成功展示了基于自然语言描述的性格特征的进化过程。
   - 未来研究方向包括详细分析当前模型、比较不同语言模型的实验、扩展游戏过程、引入不同的博弈设置以及研究人机交互的复杂社会背景下的进化场景。

通过这篇论文，作者展示了LLMs在模拟和理解人类合作行为进化中的潜力，并为未来的研究提供了新的方向。

Tencent Yuanbao

I gave Yuanbao the same prompt as the prior example. I think the person who trained Yuanbao probably did a lot of research on user reading habits, or it was probably developed by a group of people with an obsession with high-efficiency reading. Because the format it generates is clearly organized, from research background, research methods, experimental design, result analysis, to overall conclusion. It feels like taking the notes of a top student when studying. And the key data are all presented. This is what Kimi does not have under the same prompt words.

But compared to Kimi, I think the biggest difference is still with (the interpretation of) the trends in iterations of human generations (from the paper on simulations of human society). Yuanbao can tell the fluctuations of the evolving curve. In the evolutionary process, in the initial stage, after about the 300th generation, the cooperation ratio rose rapidly, reaching 0.55 around the 350th generation, and then dropped to about 0.40 around the 450th generation. Then, the cooperation ratio repeatedly increased and decreased, reaching a maximum value of about 0.75 around the 850th generation, and then rapidly dropped to about 0.15.

According to the data fluctuations, it is concluded that in the evolutionary process, the distribution of AI human personality genes in two-dimensional space shows multiple changes, reflecting the alternation of cooperative and selfish personality traits. That is, the evolution of AI humans has been jumping back and forth between selfishness and cooperation, and it also provided specific time periods. (History is really a wheel~)

Moreover, I also found that there is an extra button in the lower left corner - in-depth reading of the document. Once I clicked in – I will kowtow to you today Master Yuanbao. Please don’t abandon me from now on and take me with you.

Because it directly combines the charts and contents, turning the paper into courseware. In the past, I was annoyed when I opened the paper and saw the charts, because I had to read the small words to understand what the charts described. Now I use Yuanbao to open the charts, and I am blown away, because I understand it directly.

And I wonder if Tencent has invited a gold medal lecturer to prepare lessons. The visual design of the entire user interface is very consistent with reading habits. There is an outline of the paper on the left, and the main text is combined with the pictures to read the paper. If you don’t understand, you can also ask questions about the content in real time. It really understands me.

At the end, they also list key questions and answers. This thing shocked me. My friends, those who have participated in defenses should know the value of this function, right? This is Professor Yuanbao simulating the graduation defense with you. The teacher is marking the key points for you before the exam, and you can also refresh different questions.

They will even evaluate the paper. In other words, if you upload your own paper to Yuanbao, Yuanbao will teach you how to revise the paper, and after that, they will also simulate the defense with you. Brother (Yuan)Bao is not only good at reading papers, but I found that writing papers and simulating defenses are also very effective.

Tongyi Qianwen

The overall thought process looks good. The beginning introduces the research focus of the paper concisely and clearly, and the main text shows the characteristics and results of the research. However, if you delve into the specific content, you will find that it is not very comprehensive and a little vague. Reading the content of a conversation is better than conversing.

Claude-3.5

At first glance, Claude's reply is really concise. It mainly summarizes some key points of the paper. It is not particularly systematic, but I have to say that I actually read it because of the small number of words. But it is too concise. After reading it, I have no more to follow-up on. It is not very friendly for me, a beginner.

Of course, Tongyi Qianwen and Claude-3.5 also summarized the specific figures in the content, as Yuanbao did. The difference is that Claude-3.5 can clearly know which picture the specific conclusion corresponds to, which Tongyi Qianwen does not have. But Claude-3.5 does not put the picture there like Yuanbao does, and you have to flip the picture and slide it back and forth, which looks very troublesome.

From the test of kimi, Tongyi Qianwen, Tencent Yuanbao and Claude-3.5, I unexpectedly found that the interaction design of kimi and Tencent Yuanbao is very smooth. When asking questions and getting corresponding feedback, these two companies have one very nice thing. Click the share icon in the lower right corner of the answer generation, and they can quickly generate a long picture or link of the content.

In fact, Tongyi Qianwen will also have corresponding interaction when clicking on the share, but currently it can only copy the link of the answer, and there is no function to generate pictures. Tongyi, ah, you can improve this.

In addition to the ability to summarize papers, I don’t know how each company performs in reading research reports. Let’s try it again and see the effect.

3. Analyze research reports

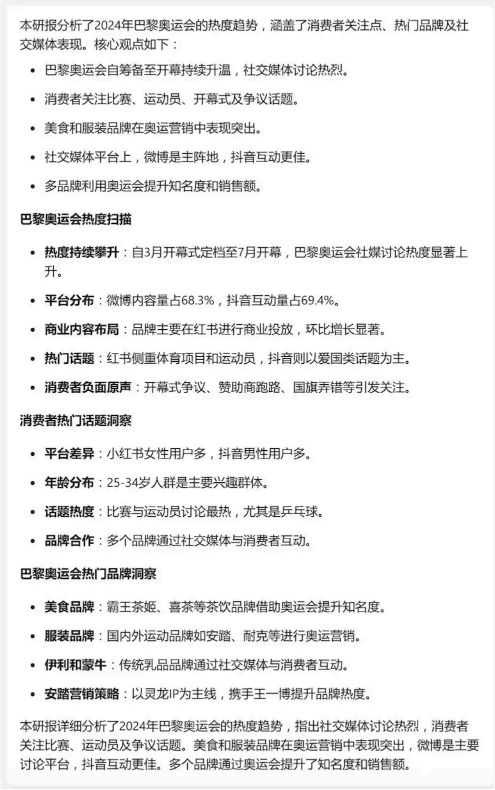
Next, I threw a PDF of "2024 Paris Olympic Games Hot Trend Insights" and added a prompt to help me analyze this research report and summarize the most important information. The number of words should not exceed 500.

Tongyi Qianwen

It is a very simple summary of a paragraph. Looking closely at the content, it only covers collaborations between platforms and brands, and the summary is not comprehensive.

Tencent Yuanbao

Here is Yuanbao again, summarizing the core points of the research report, and summarizing the specific content from the Olympic Games heat scan, topic insights, and brand insights. It is very clear.



本研报分析了2024年巴黎奥运会的热度趋势，涵盖了消费者关注点、热门品牌及社交媒体表现。核心观点如下：

- 巴黎奥运会自筹备至开幕持续升温，社交媒体讨论热烈。
- 消费者关注比赛、运动员、开幕式及争议话题。
- 美食和服装品牌在奥运营销中表现突出。
- 社交媒体平台上，微博是主阵地，抖音互动更佳。
- 多品牌利用奥运会提升知名度和销售额。

巴黎奥运会热度扫描

- **热度持续攀升**：自3月开幕式定档至7月开幕，巴黎奥运会社媒讨论热度显著上升。
- **平台分布**：微博内容量占68.3%，抖音互动量占69.4%。
- **商业内容布局**：品牌主要在红书进行商业投放，环比增长显著。
- **热门话题**：红书侧重体育项目和运动员，抖音则以爱国类话题为主。
- **消费者负面原声**：开幕式争议、赞助商跑路、国旗弄错等引发关注。

消费者热门话题洞察

- **平台差异**：小红书女性用户多，抖音男性用户多。
- **年龄分布**：25-34岁人群是主要兴趣群体。
- **话题热度**：比赛与运动员讨论最热，尤其是乒乓球。
- **品牌合作**：多个品牌通过社交媒体与消费者互动。

巴黎奥运会热门品牌洞察

- **美食品牌**：霸王茶姬、喜茶等茶饮品牌借助奥运会提升知名度。
- **服装品牌**：国内外运动品牌如安踏、耐克等进行奥运营销。
- **伊利和蒙牛**：传统乳品品牌通过社交媒体与消费者互动。
- **安踏营销策略**：以灵龙IP为主线，携手王一博提升品牌热度。

本研报详细分析了2024年巴黎奥运会的热度趋势，指出社交媒体讨论热烈，消费者关注比赛、运动员及争议话题。美食和服装品牌在奥运营销中表现突出，微博是主要讨论平台，抖音互动更佳。多个品牌通过奥运会提升了知名度和销售额。

If you are a short video operator or merchant, you will find how valuable Yuanbao's information is. First, he will tell you what the main hot spots are. Then he pointed out the two most popular social platforms, Weibo and Douyin, among which Weibo's content volume accounts for 68.3%

of the entire web, and Douyin's interactive Olympic topic interaction volume accounts for 69.4% of the entire web.

But Yuanbao also pointed out that brands mainly conduct commercial placement on Xiaohongshu because Xiaohongshu's hot topics focus more on sports and athletes, while Douyin focuses on patriotic topics. At the same time, from the perspective of consumer trends, Xiaohongshu has more female users, Douyin has more male users, and the main population is 25 to 34 years old. Doesn’t that immediately make the consumer portrait become clear? If every research report can be summarized like this, I can read 100 reports a day.

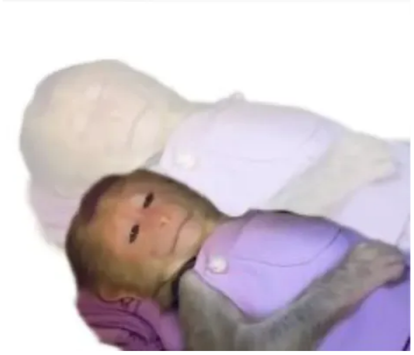
The key point is that its in-depth reading can still summarize the key information with pictures. At the end of each intensive reading, there are still a wave of answers to key questions.

Claude-3.5

Claude’s output is quite satisfactory and concisely summarizes some information you want to see. Overall, Yuanbao is indeed better at intensive reading of long articles. It is very “online” in terms of content and text format. I feel that it understands users' reading habits very well. The outline of the in-depth reading mode, the combination of pictures and texts, and the ability to ask questions about the article in real time make it very comfortable to use!

4. Extra Test

Of course, it is also very popular on the Internet recently to test the ability of AI to understand memes and reason with mathematical logic, so here we also test some of the things that everyone likes to test on the Internet to see how each company performs.



Upload this meme package and ask: What does this meme actually mean?
Tongyi Qianwen

It can be seen that it tried to understand this meme in a serious way. The physical level is there, but it has missed the chemical reaction. The humor and ennui are the key points.

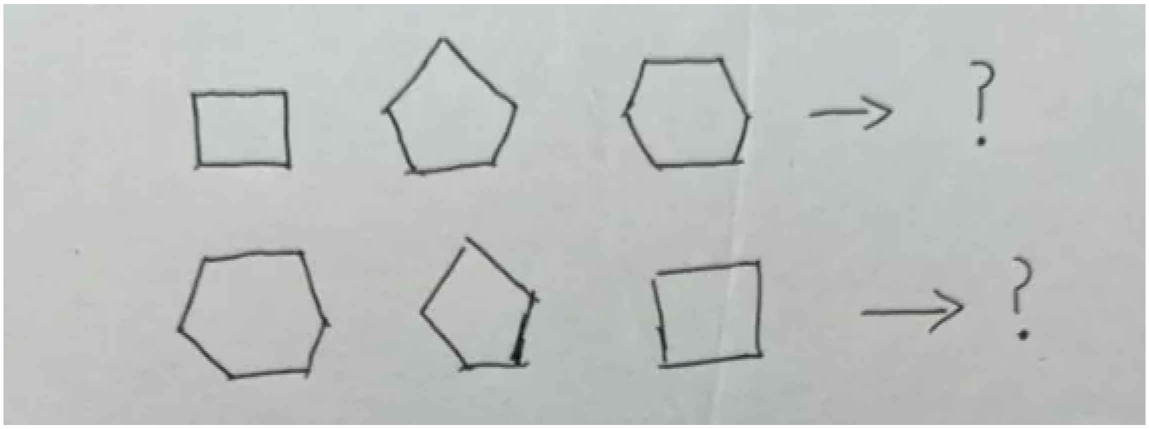
Tencent Yuanbao

Yuanbao really understands workers. It directly and clearly pinpointed the mood.
"Complaining about an unsolvable problem" or "Feeling helpless about a situation."

Claude3.5

This time, Claude enumerated many complex emotions. It seems that it is better at describing the helplessness of daily life than me.

The next is simple mathematical logic reasoning. In order to prevent the questions from being trained by AI, I will test the same figure in reverse order.



Baidu Erniebot

Wenxin Yiyan has revealed its chicken feet. The forward answer (top row) is fine, but Erniebot’s answer to the inverse direction is: simpler than the square or similar to the square…

Tongyi Qianwen passed the level normally.
Yuanbao also passed the level normally.

As an aside, when I was using Tencent Yuanbao today, I also wanted to see its ability to update the latest information in real time. The reason is that although most AIs now have the function of connecting to the Internet, they generally search for some old news as a reference.

When I tried to search for AI applications on Yiwu (online wholesale market), I actually found the article I wrote last Friday, and Yuanbao also summarized the content of the article. I tried other companies in turn, and only Yuanbao could search out (the article) at present.

In this horizontal test, there is a feeling that the big models of various companies seem to have become a little slack after last year's 100-model war. In fact, as a user, I would like to see the various companies fight back and forth, so that there will be better products to help me "work".

To be honest, the advantage of AI products lies in the process of continuous evolution. There is no eternal winner, only eternal innovators.

This is a long competition, and better user experience is the only unchanging law.