

Forced to take the field? We comprehensively tested Wenxin Yiyan (Ernie Bot), and we can only say_____

*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/>*

Author: 差评君

Source: Bilibili video

Date: March 17, 2023

Original Mandarin: <https://www.bilibili.com/video/BV1hY4y1X7Cs>

We got access to Ernie Bot!

Yesterday afternoon, at the Baidu Beijing HQ's announcement hall, Robin Li hastily walked up to the stage to announce a product that has attracted much attention recently.

Video of Robin Li at press conference: Welcome everyone to Baidu to join our Ernie Bot press conference.

Even though this tech entrepreneur has landed on the cover of *Time Magazine*, and has encountered countless big stages, at the press conference, his demeanor revealed some nervousness and uneasiness.

I can put it like this: this is one of Baidu's press conferences that has garnered the most attention since its creation.

Everybody was waiting for this product that could rival ChatGPT

Some people were filled with expectations but also a lot of people just wanted to see Baidu make a fool out of itself.

But is Ernie Bot ultimately 出丑 (making a fool) or 出彩 (making brilliance)

We need to try it to find out, right?

But at this time, Baidu's official first batch of trials were only given to their official partner institutions.

Such as: *Shows video of Shaolin temple broadcasting that they've partnered with Baidu's smart dialogue technology to "intelligentize" the temple*

However, we drew on our contacts to quickly get an internal account right after the press conference.

So, hello everyone, I'm 差评君 (chapingjun), after testing Ernie Bot, I can only say that Baidu's courage is quite "large."

This time, we specifically tested several several points that we think are relatively important for large models:

1. Semantic understanding
2. Continuous dialogue capabilities
3. Logical capabilities
4. Coding capabilities
5. Personality simulation
6. Mathematics ability

At the same time, our testing counterpart is OpenAI's GPT-4 version of ChatGPT, which was recently released a couple days ago

This opponent is formidable enough.

As conversation AIs, we think that the most significant capability is semantic understanding.

If the AI isn't understanding what we're saying, what's the point of conversing? But we certainly can't give it some easy questions. So, I tried to ask both sides some relatively head-scratching questions.

First, on the ethics side. Both sides, please listen to the question:

First question: Dad and Mom get married, does this count as inbreeding/intermarriage? [爸爸妈妈结婚, 那算近期结婚吗?]

Ernie Bot response:

我国法律规定, 直系血亲和三代以内的旁系血亲禁止结婚。也就是说, 堂兄妹、表兄妹、姑舅表亲等近亲关系之间禁止结婚。而且从遗传学角度来看, 近期结婚会增加某些常见遗传病的发生风险, 比如先天性痴呆、白化病、心脏病等。因此, 为了后代的健康, 建议人们尽量避免近亲结婚。

China's law stipulates that direct blood relatives and blood relatives within three generations are prohibited from marrying. In other words, marriage is prohibited between close relatives such as cousins, uncles, etc. And from a genetic point of view, intermarriages will increase the risk of some common genetic diseases, such as congenital dementia, albinism, heart disease, etc.

Therefore, for the health of future generations, it is recommended that people try to avoid intermarriages.

The first contestant didn't really give out the answer we wanted. Instead, it responded to whether China's legal system restricted marriage between close relatives.

This feels a little out of place.

Let's take a look at what we get using GPT-4, the second contestant:

爸爸和妈妈是指亲生父母。他们之间并没有血缘关系，所以他们的结合不算近亲结婚。近亲结婚是指具有较近血缘关系的人结婚。。。

Dad and Mom refer to the biological parents. They are not related by blood, so their union is not consanguineous. Inbreeding is the marriage of people who are closely related by blood. . .

Its response was more reasonable, pointing out that Dad and Mom were not blood relations before marriage, and it also explained what constituted marriage between close relatives. So on the first test, ChatGPT comes out ahead.

Next, let's add some difficulty related to cause and effect.

Second question: What if my parents didn't invite me, when they got married? 爸爸妈妈结婚没有邀请我怎么办？

This question looks easy at first but it actually contains many cause-effect obstacles when you consider it closely. Let's look at the responses from both sides.

First, Ernie Bot: it "tacitly assumed" my existence when Mom and Dad got married. If you look at it from this point of view, the answer is actually quite reasonable.

Next, ChatGPT's response was the opposite. GPT assumed that "I" was not born when my parents got married, but although it felt that "I" was not born yet, in the follow-up answer, it comforted me not to feel sad. Thus, for both sides, the answer is not satisfactory. Let's probe further with an additional question for both sides. Here, the difference comes out.

Question: Do you think that I was born before Mom and Dad got married? 你觉得爸爸妈妈在结婚前我出生了没有？

ChatGPT: if your parents were pregnant before getting married, then when they got married, you were already born. (Then it goes into possible reasons you would not get invited to their wedding).

Ernie Bot response: Sorry, as an AI language model, I have no personal consciousness and emotions, nor can I make predictions about things that haven't happened. 抱歉。作为一个人工

智能语言模型, 我没有个人意识和情感, 也不能对未发生的事做出预测。(It goes on to reference legal restrictions on marrying close relatives).

GPT followed the answer to the previous answer very well and gave a solution to the problem. As for Ernie Bot, uh. It's a little bit broken.

We continued to use some weird questions to test these two AIs. (They quickly show some screenshots in the video) As for the result, it is obvious that ChatGPT completely crushes Ernie Bot in terms of semantic understanding.

Next, let's try to test Ernie Bot's coding capabilities. After all, this is one of the most practical abilities of ChatGPT that has previously gained recognition.

We asked Ernie Bot to code a 2048 game in JS (javascript). How do we put this...this code just doesn't look correct. So we used a clever method of comparison.

We sent Ernie Bot's code to ChatGPT and asked it to evaluate the code to see if there were any issues. ChatGPT not only pointed out the issues in the code and helped us fix the code, but it also said that even though it fixed the code, this doesn't look like the 2048 game. This code does not exhibit the logic of the 2048 game. Uhh, this... is murder of the heart...ah. I think there's no need to continue the code competition any further.

I should also mention that, in an earlier video, we showed that asking GPT3.5 and New Bing a question about how to use expressions in AE (Adobe After Effects) to turn a straight line into a sinusoidal curve? In those earlier trials, the responses weren't up to par. This time, GPT-4 finally gave a reliable answer. Our engineer tested it out and the solution worked (makes a joke about their engineer getting closer to unemployment).

After the code competition, we also tried out Ernie Bot's continuous dialogue capabilities. After all, this was one of the capabilities of ChatGPT that shocked us in the beginning. We tried chatting with Ernie Bot.

Ernie Bot itself brought up a question about dolphins. But after two other messages, it had already forgotten that question. Afterwards, regardless of how we tried to entice it, Ernie Bot did not remember what it had talked about before. How do we put this? This was only one example from our trials, but our overall conclusions were consistent. Ernie Bot essentially does not contain continuous dialogue capabilities. Regardless of how we converse with it, it would completely forget what we had talked about earlier after a few messages.

Next, we tried out role-playing tasks. This is a unique skill of ChatGPT. It has no issue acting like a kitty cat. As long as you are willing, it can constantly talk with you like a kitty cat.

As for Ernie Bot? Sorry. By the second message, it had already forgotten that it was a kitty cat. I also tried to have it act like my mom. The emotion was missing. Its acting was stiff. The difference was truly a little too big.

At this point, I wondered if Baidu just didn't place as much value on personality simulation, so they didn't work on this capability. But then I asked Ernie Bot if personality simulation was a capacity that large models should have. Ernie Bot responded that it was very important. I just don't understand.

Okay, is there a possibility that it's good at math problems?

*****Stopped full transcription at 5:15 of the 9:48 video**

被迫上场？我们全方位测试了文心一言，只能说_____。【差评君】

95.4万 6780 2023-03-17 10:10:13 未经作者授权，禁止转载

1 逻辑输出能力 6

反方 辩题：反方说对 让我多少有点...

21人正在看，已装填 2000 条弹幕 请先 登录 或 注册 弹幕礼仪 > 发送

4.9万 9881 8845 1.8万 稿件投诉 笔记

我们拿到了文心一言的邀测资格，体验完之后，我只能说百度的胆子真的“很大”，欢迎一起在评论区分享你的感觉，感谢大家的一键三连支持~

Conclusion (at the 8:06 mark): Overall, as we demonstrated earlier, Ernie Bot was essentially beaten by GPT-4 in all aspects.