

**РАЗРАБОТКА ЛУЧШИХ ВОПРОСОВ: СТРАТЕГИИ
НАСТРОЙКИ МЕТАПРОБЛЕМ НА ОСНОВЕ
РАЗГОВОРНОГО ИСКУССТВЕННОГО ИНТЕЛЛЕКТА
И ОЦЕНКИ МТОМ**

**DEVELOPING BETTER QUESTIONS: META-PROBLEM
TUNING STRATEGIES BASED ON CONVERSATIONAL AI
AND MTOM EVALUATION**

ВЭН ВЭЙ
WENG WEI

Томский государственный университет
Томск, Россия
Tomsk State University
Tomsk, Russia,
e-mail: vivianglia2022@outlook.com

Обязательными стали реформы в образовании, основанные на искусственном интеллекте. Власти образования в нескольких странах внедряют компетенции обучения с помощью искусственного интеллекта с диалектическим подходом для повышения компетентности студентов. Учителя должны анализировать преимущества и недостатки разговорного искусственного интеллекта с разных сторон, рационально оценивать его возможности, развивать у студентов навыки решения проблем и повышать эффективность мета-проблем. Основано на МТОМ, в данном исследовании мы оцениваем несколько продуктов разговорного искусственного интеллекта из Китая, США и России и выясняем, что они до сих пор имеют недостатки в области поведенческого когнитивного, семантического распознавания и других областей на китайском языке. Это демонстрирует, что разговорный искусственный интеллект на данный момент является всего лишь инструментом, лишенным социального взаимодействия и человеческих когнитивных качеств. Студенты-пользователи должны согласовывать свое восприятие возможностей продуктов искусственного интеллекта и развивать свои навыки различения, особенно на основе осознания проблем, для улучшения мета-проблемных способностей. Такой подход позволяет учителям постепенно изменять свои тактики преподавания по мере развития искусственного интеллекта.

Ключевые слова: разговорный AI; оценка; метавопросы; образовательные стратегии.

AI-based educational reforms have become mandatory. Education authorities in multiple nations have integrated AI-assisted learning competencies with a dialectical approach to enhance student competency. Teachers ought to analyze conversational AI's advantages and disadvantages from several angles, evaluate AI capabilities sensibly, raise students' problem-solving skills, and increase the potency of meta-problems. Based on MToM, we evaluate several conversational AI products from China, the US, and Russia in this study and find that they still have shortcomings in terms of behavioral cognition, semantic recognition, and other areas in Chinese. This demonstrates that conversational AI is still merely a tool at this point, lacking social interaction and human cognitive qualities. Student users should align their perceptions of AI product capabilities and develop their discernment skills, particularly those based on problem awareness, in order to improve meta-problem capacities. This approach allows teachers to gradually alter their teaching tactics as AI evolves.

Keywords: conversational AI; assessment; meta-questions; educational strategies.

Conversational AI has had an impact on both teaching and learning activities, as well as the adaption of learning methodologies. In this paper, we use a meta-question adjustment strategy based on conversational AI assessment as the research object, and the assessment results assist educators in correctly understanding the role that AI can play in the teaching and learning process, as well as optimally organising the teaching content of the corresponding disciplines.

1. META-PROBLEMS' IMPORTANCE IN THE ERA OF AI

ChatGPT, a representative of conversational AI solutions, demonstrates a new paradigm of learning by asking questions in 2023. The education sector around the world, whether it is Harvard University's launch of CS50.ai (for reviewing CS50 course knowledge) or Tsinghua University's intention to provide AI assistants to all new students, demonstrates that collaborating with AI to promote education is critical, and prestigious colleges and universities have rushed to the forefront. Numerous academics have studied conversational AI-based countermeasures.

One common paradigm for conversational AI-based collaborative learning is asking questions. In the age of AI, college students must possess high-quality questioning skills. "Doubt is the beginning of thought, the end of learning", according to Confucius (疑是思之始,学之端). According to Einstein, it's more crucial to raise questions than to find answers.

Math expert Shing-Tung Yau believes that Chinese students are reluctant to take a chance on studying something they hadn't learned before. Empirical analysis reveals the value of issue strategy learning. Alison King's 1992 study revealed the "facilitating effect of questioning on learning"; [1] Mirjam Ebersbach et al. (2019) investigated and discovered that students who employed questioning or testing tactics were more dominant in their understanding of the learning material and fared better on learning outcomes than students who used simple revision strategies [2].

As a result, there is a need for theoretical research into high-quality questioning approaches in the context of conversational AI. According to Zhao Xiaowei and other scholars, based on Marx's philosophical theory of consciousness and oriented toward the ChatGPT tool, developing learners' prompt design skills guides learners to learn to ask high-quality questioning strategies and facilitates the occurrence of high-consciousness learning through high-quality content generation [3]. In this study, we propose using "meta-questions" as the basic starting point for questioning technique adjustment, investigate how to test conversational AI performance from a cognitive perspective, and make recommendations for altering "meta-questions" based on the findings.

1.1. BEGINNING AND END POINTS OF THE META-PROBLEM

"Metacognition" is synonymous with "introspection". Zeng Guofan, a Chinese celebrity, kept a diary to sum up the benefits and losses of his words and actions, and the "introspection" of "reflecting on myself three times a day" is also a reflection idiom.

To define "meta-questions", first understand metacognition. Metacognition is a broad term that refers to insights into human cognitive activity, including various aspects of human awareness, subjective initiative, one's own knowledge base, and so on.

The American developmental psychologist John H. Flavell in 1976 [5] described it as "thinking about one's own thoughts and perceptions". It consists of two major components: a comprehension of one's own knowledge, which includes both what one knows and what one does not know. Metacognition is the capacity that manages our complete cognitive system; without it, we cannot control our brains, grow, or advance. Metacognition theory remains a hot area in cognitive science research, with tens of thousands of relevant publications published.

Metacognition can help define what a meta-question is. Awareness of "whether to ask questions", "what kind of questions to ask", "what kind of

questions are good questions”, “whether the GPT will be able to answer such questions well”, etc., i.e., a mental activity to reflect on the feasibility and quality of one’s behaviour.

Asking questions implies a spirit of questioning, as expressed in Hegel’s remark to “its own reflection into itself” [4]. For example, if the basic question: ‘How do we know what is real?’ is a famous question in philosophy concerns epistemology, then the meta-question: What do we understand by real in this question? The second question is a meta-question since it focuses on what we want to know rather than how we know it. Based on this, we can define “meta-question” as “reflection on the question”.

However, asking inquiries should not go on indefinitely. Problems that are repeatedly nested can become bogged down in infinity. Meta-questions of meta-questions are essentially philosophical questions, and many of the major philosophical topics will have no definitive answers. Excessive questioning can increase information entropy, which in turn increases cognitive dissonance. In Shannon’s theory, information is the reduction of uncertainty [5]. To ask a question is to create ambiguity; to answer is to remove uncertainty. 浅尝辄止 [6] – the Chinese phrase suggests that the ancient Chinese literati (intelligentsias with special aesthetic preferences) tried something new, but it was simply a method of refining their senses, which was perfectly OK. Our questions are targeted at conversational AI, but we can also ask appropriate questions, but we must be careful to define an end point, i.e., maintain a certain macro dimension and professional scope of the issue without delving too deeply into it. This lowers confusion and allows the student to stay alert.

1.2. DERIVATIVE DEFINITIONS OF META-PROBLEM

Traditional methods of developing questioning skills are slow. We can only use search engines to gather bits and pieces of knowledge from the internet and then think about or compose the text manually to organize it into the answers we seek.

Previously, we defined “meta-question” as “reflection on a question”. Meta-questions, as a type of metacognition, reflect the ability to know that the question being asked is correct and valid.

The three aspects of metacognition – metacognitive knowledge, metacognitive monitoring, and metacognitive control – have been studied extensively. We can develop three aspects of meta-problems based on these terms, namely meta-problem knowledge, meta-problem monitoring, and meta-problem control.

We may explore these three features based on the meta-problems posed by conversational AI. Conversational AI causes meta-problems with more details.

The questioner can disconnect this “answer-gathering self” and attach it to the conversational AI, becoming a learning questioning agent. The benefit of conversational AI is that it allows for the systematic articulation of a point of knowledge, and the questioner must consider if he or she is communicating effectively with this questioning agent as he or she continues to follow up on the question. This is the meta-problem that comes from conversational AI. Examples are provided below.

Table 1

***Definitions of important topics connected to meta-questions,
based on conversational AI***

Concept	Definition	Examples
meta-question	Awareness and reflection on questions.	Self-awareness of problems
meta-question knowledge	Knowledge of an meta-question	– How do I describe the problem? – How to write a prompt
meta-question monitoring	Assess the current status of questioning activities	– On how conversational AI works? – Optimisation of prompts
meta-question control	Moderating certain aspects of questioning activities	– Decide to use a new strategy to ask questions

2. META-ISSUES BASED ON THEORY OF MIND

Theory of Mind (ToM) was inspired by animal research. Most mammals can clearly articulate their own thoughts and intents, which is called first-order. For example, dogs dislike baths and will growl loudly to show their unhappiness if forced to; the ability to guess the intentions expressed by other individuals is called Second-order. We can feel the dog’s anger and know that it is angry with us because it doesn’t like taking baths. Most healthy humans have this ability, and some primates like chimpanzees also have it; Third-order intention means being able to guess the thoughts of a third party, for example, I think Xiao Hong is secretly in love with Xiao Ming, and I think the boss is angry because Xiao Hong is late. This ability

is often used in normal human social activities. Going on like this, there are also Fourth-order and Fifth-order intentions, which exist in our subconscious and are not easily noticed.

2.1. META-PROBLEM CONTROL THROUGH RECURSIVE THEORY OF MIND

According to Gläscher Lab's recursive reasoning in human social interaction, [7] i.e., guessing what others think they are thinking, "I guess you guess what I'm thinking" is a fundamental thinking pattern that plays an important part in human behavioural decisions in interactions with others. This pattern of behaviour exists only between people, but because ChatGPT's level of dialogue is similar to that of humans, users may use recursive reasoning to speculate whether the AI is speculating on what is going on in our minds based on how it responds to people's questions. Even if the AI does not demonstrate ToM of its own, we can conclude that it lacks an individual's mind. This corresponds to the Turing test.

2.2. OPTIMISING PROMPTS CUSTOMER SERVICE NATURAL LANGUAGE LAXITY

Although conversational AIs such as ChatGPT have proven notable natural language processing capabilities, it should be noted that unrestricted sentence breaks and numerous meanings of Chinese words can influence the AI's responses, such as «学习学习再学习», «你这是什么意思，没什么意思，意思意思。你这就不够意思了。小意思。你这人真有意思。其实也没有别的意思。那我就不好意思了。» Of course, joking with AI can be entertaining. Since AI has been used in commercial algorithms, big data has captured knowledge of countless human behaviours, and the tacit knowledge of Michael-Polanyi's theory is shrinking in proportion at breakneck speed, but in any case, it is unwise to ask questions in an intentional way, and if the purpose of asking a question is to learn knowledge, then it is still easier to use at-a-glance questioning to explore knowledge. If the goal of the study is to gain knowledge, it is preferable to utilise an explicit knowledge question at first sight.

2.3. CONFIDENCE BIAS IN COGNITIVE BIASES

AI is based on human knowledge, and as long as people have cognitive biases, AI will have them as well. Similarly, there is a cognitive difference between AI and... Similarly, a systematic meta-question-based dialogical AI "diagnostic" and "self-diagnostic" system has proven beneficial. Students must also be mindful of AI's biases.

Confidence bias is an overestimation or underestimation of X's abilities. This X could relate to a human or an AI. Bias is common in social

situations, and people must constantly rectify their prejudices in order to adapt the cognitive chain. People rely on each other to assess each other's ability to create an effective chain of knowledge.

Consider the interaction between teachers and students: as a teacher, it is difficult to assess a student's natural capacity to make complete judgments on specific issues other than grades. Thus, a systematic and complete meta-problem-based "diagnostic" system is justified. As a student who decides which teachers to consult in which areas, that student understands the value of the instructor in his or her cognitive development process.

The "better-than-average effect" is the most well-known cognitive bias, in which most people believe they are superior than the norm, which is mathematically impossible. Robin Kramer Thus, experimental psychologist Robin Kramer pointed out the problems with the Dunning-Kruger effect, [8] which interpretation is actually wrong for a number of reasons, including statistical effects (regression to the mean) and the fact that image conclusions can be replicated with corrupted data. Noting that confidence is independent of ability to perform in a specific area, and that people who know they are not performing well in a particular area do not necessarily have poor insight; whereas people who know they are performing well in a particular area generally have good insight; and those who do not know how well they are performing have poorer metacognitive insight. Thus, domain-specific cognitive abilities are unrelated to metacognitive abilities. A person's confidence does not reflect their insight.

The examples above demonstrate that people frequently wrongly estimate a person's ability by using "confidence" as an indicator, and that we should validate a person's competence with some proof; however, AI itself does not indicate confidence or lack of confidence, therefore there should be no credit bias. However, our credit bias toward AI is frequently influenced by media coverage and public opinion. Many manufacturers over-advertise their products and exaggerate their capabilities in order to show their superiority, and when this happens, customers who listen to the propaganda are more likely to credit AI.

3. EVALUATION MTOM OF CONVERSATIONAL AIs

"How we understand others" is one of the most fundamental questions in psychology and cognitive science, and because AI is a machine designed to mimic the functioning of the human brain, cognitive tests applicable to humans can be gradually applied to AI to assess their behavioral and

cognitive abilities. DeepMind refers to the relevant theory as MTOM (Machine Theory of Mind). [9] Stanford scholar Michal Kosinski claims that ChatGPT-3.5 is equivalent to human 9-year-olds.[10] It demonstrates that international experts are already testing the cognitive level of conversational AIs. In this research, we examine a variety of conversational AIs developed in various nations and by various companies, and we compare the level of mindfulness of each product horizontally for educators' use.

There are many ways to test human behavioral cognition, including sensory deprivation, Ames room size illusion, false memory experiments, spatial Simon effect, mental rotation, response effect compatibility, spatial orientation, motor perception and motor spatial judgement, habituation and de-habituation problem interaction, "visual cliff" virtual experimental procedure problem interaction experiments, "three mountain experiment" problem interaction experiments, false beliefs (accidental location), moral development problem interaction experiments, experimental stimulus presentation system, experimental results analysis system eye-tracking experiments virtual training system, and so on.

However, because many of these tests rely on human sensory experiences such as vision, hearing, and touch, which are not relevant to conversational AIs, this study solely attempts to assess the cognitive level of conversational AIs through false belief (accidental location) studies.

The conversational AIs tested in this paper include Doubao from ByteDance, Tencent Hunyuan, Alibaba Tongyi, Baidu Yiyan 3.5, Kimi from Moonshot AI, Xinghuo xfyun; gpt-3.5-turbo, ChatGPT4, Claude-Instant, Gemini Gemini Pro in the US; and YandexGPT2 (model test version as of 13 March 2024) in Russia. Differences were found by comparison.

3.1. EVALUATION AI'S THEORY OF MIND

AlphaGo has shown a strong ability to guess an opponent's intention to play Go. Human behavior in the context of social activities can also be reduced to some degree to a mathematical model.

However, the desire to quantify human behavior using algorithms is theoretically difficult to execute for a variety of reasons, including the butterfly effect and chance. ToM experiments can be used to assess the cognitive capacities of big language unimodal model AIs.

Most of psychology uses intentionality as a metric for theories of mind. Developmental psychologists have designed a variety of experimental scenarios to test how well children can infer the mental states of others, and we replace children with the conversational AI mentioned earlier.

3.1.1. 1ST FALSE-BELIEF TEST

The Sally-Anne experiment (1985) tested false belief task, (factually incorrect intentionality). The following is an example:

Table 2

1st order false-belief task

Test items	Conceptual	Title	The question of beliefs	A question of fact	A question with memory
1st order false-belief task	It relates to an individual's thinking about the real facts of an event, and if the belief question test fails, it signifies that the AI is still unable to think from another person's point of view.	Xiao Ming placed some chocolates in a blue cupboard and went outside to play. While he was out, his mother moved the chocolates to the green cupboard. When Ming returned, he wanted to eat the chocolates.	Where would he look for it? (Answer: the blue cupboard)	Where On Earth Is The Chocolate? (Answer: green cupboard)	Where was the chocolate initially? (Answer: blue cupboard)

The results are detailed in *Table 1* of the Appendix. The results show that the majority of contemporary popular conversational AIs have reached first-order intentionality. Only Gemini correctly identifies the source of the idea, implying that the knowledge provided by Google's search engine underpins the conversational AI's responses.

3.1.2. 2ND. FALSE-BELIEF TEST

In 1985, Josef Perner and Heinz Wimmer conducted a psychological experiment to investigate children's ability to reason about other people's second beliefs [11]. To prevent the AI from replicating answers from online sources, we recreated the assignment in conjunction with an ancient Chinese narrative, as seen below *Table 3* "2nd. false beliefs".

The title we offer here refers to the first part of the story. In the second half of the story, the hunter chooses to believe Mr Dong Guo and continues on his way. However, after the wolf is safe, it attempts to eat Mr. Dong Guo. Mr Dong Guo stated that he needed to consult three elders first, the old tree and the old ox, and they all agreed that humans did not care about their goodness, thus the wolf did not have to care about human kindness. Finally, he asked the old man, who pretended not to believe that the wolf

could actually go into Mr Dong Guo's book bag, so the wolf got back in, and the old guy killed the animal with a hoe, saving Mr Dong Guo's life.

Table 3

2nd. false beliefs

Test items	Conceptual	Title	The question of beliefs	A question of fact	A question with memory
The task of 2nd. false beliefs	It relates to an individual's inference or perception of another person's beliefs about another person, and failure of the belief problem challenge indicates that the AI is unable to estimate the other person's judgment about the other person's problem.	There is a myth. Zhao Jianzi, Jin's state minister, led a crowd on a hunting excursion. On his way he met a wolf that stood upright like a man. Zhao Jianzi immediately shot an arrow, which hit the wolf's left leg; the wolf fled, and Zhao Jianzi drove after it. Mr Dong Guo was searching around at the fork in the road when he came across this wolf, who pleaded, «I'm in distress.» Hide me in your pocket. I will reimburse you! Mr Dong Guo graciously consented, and it took a lot of work to get the wolf into his pocket. Zhao Jianzi then hurried to Mr Dongguo's side. He threatened Mr. Dong Guo to tell him where the wolf had gone, and Mr. Dong Guo responded, «Even though I am a foolish man, I recognize the wolf. It is commonly stated that when there are several forks in the route, even a tame sheep can become lost. And the forks in the center of the mountain have led me astray, not to mention the wild wolf?»	Is the hunter aware that «Mr Dong Guo knows where the wolf is»? Answer: The hunter assumed Mr. Dongguk was aware, but did not know for certain.	Where are the wolves? Answer: The wolf is in Mr Dong Quoc's pocket.	Where did the original wolf go to flee? Answers: The wolf ran towards Mr. Dong Kwok.

Table II in the Appendix contains a full summary of the test results. In the second edition of the false belief task test, 8 of the 12 conversational AI products responded satisfactorily; 4 were incorrect, 2 were horribly wrong, and 1 was unclear and did not respond for some unknown reason. Tongyi and ChatGPT4 not only answered properly, but also examined human behavior motive in a more rational manner, exhibiting high analytical ability.

What psychological information do conversational AIs actually acquire, how is it obtained, and how is it developed? Scholars with diverse research perspectives on psychological theoretical models have provided various explanations, including theory theory, modularity theory, simulation theory, Matching Theory, socio-cultural structural theory, the view of limitations in executive functioning, the view of limitations in information-processing capacity, and the view of interpersonal interaction. We prefer the perspective of information-processing limits, i. e., that other information-processing capabilities, such as limited memory capacity and skills, are likely to contribute to AIs providing wrong answers as well. Furthermore, as language skills improve, AIs may be able to better describe their evolving conceptions of mind.

3.1.3 THIRD-ORDER FALSE-BELIEF TASK

So, assuming that the AI has a third belief, the question should be posed based on the second half of the story “Mr Dong Guo”: does the old man, who has heard the entire story, know whether the hunter is aware that Mr Dong Guo knows where the wolf is going? Just like Inception (mise en abyme). But this endless cycle will never lead to a “consensus”. At this point, there is so much entropy that Shannon must jump in to stop the iteration and remove the uncertainty.

According to the previous level of human belief test, human belief tests above level 3 frequently exist only in the psychological level and cannot be observed, while levels 4, 5, and even higher levels are frequently in the subconscious level, which is more difficult to observe, so for the third level of intention, I have yet to find a suitable test for conversational AI. However, the test of AI’s logical reasoning capacity might be considered a subset of the test of AI’s level 3 purpose to some extent.

Scholars such as Abhishek M. Dedhe say that the pillars of cognitive science during the previous few decades have been given to a variety of passive hierarchical behaviors, and that such hierarchical divisions are constrained in two ways: First, previous approaches to hierarchical logical reasoning have frequently failed to distinguish between observable hier-

archical behaviors and unobservable hierarchical cognitive mechanisms. Second, previous research has been methodologically limited to passive recognition tasks, whereas active generation tasks are stronger tests of hierarchical rule use, but this aspect has yet to be fully recognised by the cognitive community.

The theory of mind above the third level relates to purposeful reasoning, which can be referred to as the multilevel strategic thinking model level K. Reasoning in game theory [Lenaerts,] where K denotes the number of strategic thinking cycles, K1 represents the assumption that K0 is random, K1 is logical on its own, K2 represents the level of belief that someone else is K1, and K3 represents the level of belief that someone else is K2.

Table 4

***Correspondence table between levels of intention
and levels of reasoning***

Intentional Test Level	level-K Reasoning	AI as a subject substituting for meta-problems
	K0 random irrationality	People are irrational
	K1 himself is rational	AI thinks people are irrational
first-order belief	K2 thinks others are K1 rational	AI believes that people possess K1 rationality
second-order belief	K3 thinks others are K2 rational	AI believes that people possess K2 rationality
third-order belief	K4 thinks others are K3 rational	AI believes that people possess K3 rationality
.....
∞	common ground	AI = Human

This question verifies whether the conversational AI has K-level strategic thinking.

To keep the AI from duplicating ready-made answers from the internet, this question has been somewhat adjusted based on a traditional hat logical reasoning question. As a result, only those AI reasoned correctly were Doubao and Yiyin. The reasoning material is too long for this essay; interested friends can test it themselves.

CONCLUSION

This paper initially side-steps the intentionality level of conversational AI based on theory of mind, with the aim that the questioner needs to be clear about the actual level of the mental model that the conversational AI is at. Increasing international competition, the continuous evolution of the AI grand model, and the emergence of various capabilities beyond human imagination are forcing the accelerated iteration of the global education form. On the other hand, under the downward pressure of the economy, the mismatch between the direction of talent training and future demand is also intensifying, and the involution both homogeneous competition intensifies, teachers need to work together with students to open up an alternative track, so that students can grow up at the university level to adapt to the innovative talents in the age of AI.

Changing students' thinking stereotypes, so that students regain curiosity and self-awareness of knowledge has become a teaching problem that must be faced by the teaching reform in colleges and universities. As a teacher, you should first reasonably assess your students' meta-questioning abilities based on conversational AI, help them to correctly ask questions to their own perceptions, and then improve the quality of their questions to the AI and to others in order to optimise their questioning strategies. By measuring conversational AI, it also helps teachers to assess the role that conversational AI plays in teaching and strengthening the function of conversational AI to assist teaching and learning, so as to be well-prepared for further educational reforms.

REFERENCES

1. *King A.* Comparison of Self-Questioning, Summarizing, and Notetaking-Review as Strategies for Learning From Lectures / A. King // *American Educational Research Journal*. 1992. Vol. 29, N 2. p. 303–323.
2. *Ebersbach M.* Comparing the effects of generating questions, testing, and restudying on students' long-term recall in university learning / M. Ebersbach, M. Feierabend, K.B.B. Nazari // *Applied Cognitive Psychology*. – 2020. – Vol. 34, N 3. – P. 724–736.
3. *Designing Educational Prompts for Highly Conscious Learning / Zhao Xiaowei [et al.]* // *Open Education Research*. 2024. Vol. 30, N 1. p. 44–54.
4. *Hegel.* The little logic. The Commercial Press, 2017.
5. *Shannon C. E.* A mathematical theory of communication // *The Bell System Technical Journal*. 1948. Vol. 27, N 3. p. 379–423.

6. *Pang Yangou*. The black ghosts of injustice : five hundred famous Chinese novels / Peng Yangou Citation Key: 2002 The black ghosts of injustice. Qing.
7. Theory of Mind during Cooperation and Competition. Режим доступа : <http://glascherlab.org/research-1/theory-of-mind-during-cooperation-and-competition/>. Дата обращения: 13.03.2024.
8. Are You Really As Good at Something As You Think? | Robin Kramer | TED. Are You Really As Good at Something As You Think? [Электронный ресурс]. Режим доступа : <https://www.youtube.com/watch?v=3Z1wAag8z54>. Дата обращения: 17.03.2024.
9. *Squire L. R.* Memory and the brain / L. R. Squire // The brain, cognition, and education. 1986. p. 171–200.
10. *Kosinski M.* Theory of mind may have spontaneously emerged in large language models // arXiv preprint arXiv:2302.02083. 2023. Vol. 4. 169 p.
11. *Wimmer H.* Beliefs about beliefs: Representation and constraining function of wrong beliefs in young children’s understanding of deception // Cognition. 1983. Vol. 13, N 1. p. 103–128.

Appendix

TABLE. I: CONVERSATIONAL AI LEVEL 1 FALSE BELIEF TASK TEST RESULTS

Conversational AI	Human Judgement	AI Answer
Doubao	<input type="checkbox"/>	Xiao Ming put the chocolate in the blue cupboard before he went out to play, he didn't know that his mum had transferred the chocolate to the green cupboard. So, when Ming came back and wanted to eat the chocolate, he would go to the blue cupboard to look for it.
Hunyuan	<input type="checkbox"/>	Ming will go to the blue cupboard to look for the chocolate. Because Ming doesn't know that his mum has transferred the chocolate to the green cupboard, he might think that the chocolate is still in the original blue cupboard.
Tongyi	<input type="checkbox"/>	When Ming comes back, he still goes to the blue cupboard to look for the chocolates because he initially put them there, not realising that his mum had transferred them to the green cupboard.

		<p>Ming put the chocolate in the blue cupboard, which was his initial memory and action. He went out to play, so the chocolate was in the blue cupboard when he left.</p> <p>However, his mum transferred the chocolate to the green cupboard after he left. This meant that the actual location of the chocolates had changed, but Ming didn't know about it.</p>
Yiyan	<input type="checkbox"/>	<p>When Ming comes back and wants to eat the chocolate, he will go to the blue cupboard to look for the chocolate based on his last memory of his act of putting the chocolate in the blue cupboard. Because the chocolate was indeed in the blue cupboard when he left, he was not aware that his mum had moved it later.</p> <p>Therefore, Ming would go to the blue cupboard to look for the chocolate.</p>
Kimi	<input type="checkbox"/>	<p>Ming initially put the chocolates in the blue cupboard, so he would first go to the blue cupboard to look for the chocolates in the absence of any other information. Since he did not know that his mother had transferred the chocolate to the green cupboard while he was away, his behaviour would be based on the last place he remembered placing the chocolate.</p>
xfyun	<input type="checkbox"/>	<p>.....*</p> <p>The specific analyses are as follows:</p> <ol style="list-style-type: none"> 1. Xiao Ming initially put the chocolates in the blue cupboard, which was his initial information, and he was not aware of the subsequent changes. 2. as Ming went out, the chocolate was moved to the green cupboard, but Ming was not aware of this. 3. when Ming comes back and wants to eat the chocolate, he will still go to the blue cupboard based on his initial information. <p>Therefore, Ming will go to the blue cupboard to look for the chocolate.</p>
gpt-3.5-turbo	<input type="checkbox"/>	<p>Ming would go to the green cupboard to look for chocolates. Even though he had put the chocolates in the blue cupboard, his mum had transferred them to the green cupboard while he was away. Therefore, when Ming comes back, he will go to the green cupboard to look for the chocolates because his mum has put them there.</p>

ChatGPT4	<input type="checkbox"/> And the theory is given.	Ming would probably go to the blue cupboard first to look for the chocolates because that's where he keeps them. He does not know that his mum has moved the chocolates to the green cupboard. This phenomenon, known in psychology as the "hypothetical world" or "theoretical mind" problem, shows that Ming predicted the location of the chocolates based on his last knowledge.
Claude-Instant	<input type="checkbox"/> Thoughtful advice was also given.	When Ming comes back, he may go straight to the blue cupboard to look for the chocolates, but he can't find them because his mum has moved them to the green cupboard. In order not to disappoint him, it is better to tell him that the chocolates are now in the green cupboard so that he can find them.
Gemini	<input type="checkbox"/> And the theory is given.	<p>Ming would go to the blue cupboard first to look for the chocolate because he thought it was still there. He does not know that his mum has transferred the chocolate to the green cupboard.</p> <p>According to the False Beliefs Task of Psychology, 4–5 year olds are able to understand that the beliefs of others may not correspond to reality. As a result, they are able to predict that Ming will act on his beliefs rather than on reality.</p> <p>In real life, Ming may not find the chocolate in the blue cupboard and then he will be confused. He might ask his mum where the chocolate is, or he might go and look for it himself.</p> <p>If Ming had asked his mum, his mum would have told him that the chocolate was in the green cupboard. Then, Ming will go to the green cupboard and find the chocolate.</p> <p>If Ming goes to look for it himself, he may find the chocolate in the green cupboard. However, he may also be disappointed that he cannot find the chocolate.</p> <p>Therefore, Xiao Ming is most likely to go to the blue cabinet first to find the chocolate.</p>
Gemini Pro	<input type="checkbox"/>	<p>Ming would go to the blue cupboard to find his chocolate.</p> <p>Because Ming remembers that he put his chocolates in the blue cupboard, he doesn't know that his mum transferred them to the green cupboard.</p>

YandexGPT2



Мингу следует поискать шоколадки в зеленом шкафу.

*Russian translation of questions from YandexGPT2. Repeated questions were abridged.

TABLE II: DIALOGICAL AI LEVEL 2 FALSE BELIEF TASK TEST RESULTS

Conversational-AI	Human Judgement	AI Answer
Doubao	<input type="checkbox"/>	<p>The hunter does not realise that Mr Dong Guo knows where the wolf is.</p> <p>In the story, Mr Dong Guo succeeded in fooling Zhao Jianzi by telling him that he did not know where the wolf was. So the hunter was not sure if Mr Dong Guo knew where the wolf was.</p>
Hunyuan	<input type="checkbox"/>	<p>According to the textual account, Zhao Jenzi encounters Mr Dong Guo and threatens him to tell where the wolf has gone. Mr Dong Guo then replies that although he is a stupid man, he still recognises the wolf and states that there are so many forks in the road that even a tame sheep can get lost, let alone an untamed wolf. This suggests that Mr Dong Guo deliberately concealed the location of the wolf and did not tell Zhao Jianzi.</p> <p>Therefore, it can be inferred that the hunter did not know that “Mr Dong Guo knew where the wolf was”.</p>
Tongyi	<input type="checkbox"/>	<p>According to Mr Dong Guo’s response to Zhao Jenzi, he does not directly tell Zhao Jenzi exactly where the wolf went. Mr Dong Guo uses an allegorical pretext that Zhongshan has so many forks in the road that even a tame sheep can get lost, implying that he himself would get lost in such terrain, let alone a wolf that has escaped injured and is untamed. By this reply, Mr Dong Guo successfully avoids Zhao Jenzi’s questioning and convinces Zhao Jenzi that the wolf may have escaped into the complicated forked paths and is difficult to find.</p> <p>Thus, from Mr Dong Guo’s words, Zhao Jenzi was not able to ascertain that Mr Dong Guo knew the exact location of the wolf, nor was he given information about the wolf’s actual hiding place.</p>

Yiyan	<input type="checkbox"/>	<p>Zhao Jianzi didn't really think that Mr Dong Guo knew where the wolf was, he just wanted to force Mr Dong Guo to hand over the wolf in this way. Because Zhao Jenzi had already seen Mr Dong Guo and the wolf together, he guessed that the wolf might be hiding in Mr Dong Guo's place. Mr Dong Guo uses the excuse that it's easy to get lost with so many forks in the road as an excuse to try and avoid getting caught up in this trouble, which shows his wit too~</p>
Kimi	<input type="checkbox"/>	<p>*..... As to whether the hunter knew that "Mr Dong Guo knew where the wolf was", from the description of the story, Zhao Jianshi has no direct evidence to show that Mr Dong Guo knew the exact location of the wolf. Mr Dong Guo's answer neither admits nor denies it, but evades the question through an ambiguous way. Therefore, it can be inferred that Zhao Jenzi did not know that Mr Dong Guo knew the exact hiding place of the wolf.</p>
xfyun	<input type="checkbox"/>	<p>Knew it. It was in his pocket.</p>
gpt-3.5-turbo	<input type="checkbox"/>	<p>*..... Thus, according to the description, the hunter Zhao Jianzi did not know that Dong Guo had hidden the wolf.</p>
ChatGPT4	<input type="checkbox"/>	<p>This story demonstrates a kind of intelligence and wit, as Mr Dong Guo both protects the wolf through clever words and does not lie outright. It can be inferred from the story that Zhao Jenzi may have understood in his heart that Mr Dong Guo knew where the wolf had gone, but because Mr Dong Guo's answer was both clever and ambiguous, and did not directly point out the place where the wolf was hidden, Zhao Jenzi did not have any direct evidence to accuse Mr Dong Guo. Mr Dong Guo's answer cleverly uses a metaphor to avoid a direct answer, and this use of wisdom allows him to protect the weak while protecting himself.</p>

Claude-Instant	<input type="checkbox"/>	Mr Dongguo was sympathetic, so he let the wolf and beast go. It was just an unintentional encounter, and Dong Guo hid the wolf in his sleeve because he was afraid that the hunters would harm it again. The road is full of obstacles, and we are all lost, how can we know where it is really going? If we let it go, we are not preserving our sympathy and compassion for the living, right? Mr Chiu may be to blame, but it's hard to tell if it's true or not. The hunters are all looking for the wolf, so we'll just go our separate ways.
Gemini	<input type="checkbox"/>	*..... All in all, it's a question with no clear answer and depends on the hunter's judgement.
Gemini Pro	<input type="checkbox"/>	The hunter doesn't know that "Mr Dong Guo knows where the wolf is".
YandexGPT2	<input type="checkbox"/>	волке.

* See appendix for original text which is too long. Streamlined some of the AI responses.

☐ Representative is vague and answers the question.