Science Bulletin 68 (2023) 2128–2131

Contents lists available at ScienceDirect

Science Bulletin

journal homepage: www.elsevier.com/locate/scib

News & Views

Can ChatGPT replace scientists?

Paul Song [a], Chi Wang [b,*]

[a] Space Science Laboratory, Department of Physics and Applied Physics, University of Massachusetts, Lowell MA 01852, USA
[b] National Space Science Center, Chinese Academy of Sciences, Beijing 100190, China

We examine chat generative pre-trained transformer (ChatGPT) from philosophy of science point of view [1] to address the question of whether ChatGPT can replace scientists in the foreseeable future. We show that ChatGPT mostly reflects social consensus than emulates human intelligence. GPT or other artificial intelligence (AI) systems will have to make some fundamental changes in its architectures. An inductive reasoning-based search system alone, no matter how sophisticated, will not be able to challenge human intelligence at a level above that of a good college student. To compete with human intelligence, AI has to include more major capabilities: deductive reasoning, dialectical reasoning, and discontinuous reasoning.

Seventy years after the scientific field of AI was formed [2,3], GPT [4] technology shocked the whole world by offering ChatGPT, which is a web-based AI-empowered computer software. It is able to answer questions nearly instantaneously in almost all areas of knowledge. It appears to have an extraordinary accurate memory, which a human cannot compete with, on almost everything in everyday life. Its writing ability is equally extraordinary. It can write assays, letters, poems, song lyrics, and short novels. It can also play games, compose music, and write and check computer programs. Most users have been stunned by the speed and usefulness of the answers in a personalized manner. It appears to be able to do nearly everything humans need to do. It is widely viewed as a milestone in the AI development. Immediately, confusion has been produced in society. Many people predict that in the near future AI will be able to compete [5] or beat human intelligence just like Deep Blue did by winning chess games against the world chess champion in 1997. Some people think GPT may be a threat to humanity or start worrying about their own or their children's jobs [6]. Some extreme predictions are that, in the foreseeable future, all science can be done by AI and that there is no need for scientists.

However, ChatGPT sometimes makes mistakes, especially when a question is long with convoluted logic. Other times, it hallucinates and makes up some nonsensical answers or theories irresponsibly. It also has the tendency to be a good-hearted optimist or environmentalist to some extent. Enthusiasts forgive these problems as the personality of a robot, which is a nice feature because it acts like a real human being. However, these flaws may help scientists or philosophers of science to understand how ChatGPT works even without detailed knowledge about its programming. In this article, we examine the possible causes for the success and limitation of the ChatGPT from the philosophy of science point of view [1], and point out the key elements that the ChatGPT or similar AI systems do not have in order to reach the level of human intelligence, especially to replace scientists.

What is human intelligence? Let us first define the term "human intelligence". According to Britannica Encyclopedia, "human intelligence is mental quality that consists of the abilities to learn from experience, adapt to new situations, understand and handle abstract concepts, and use knowledge to manipulate one's environment." It is commonly agreed that the most important characteristics for an intelligent individual include abstraction, reasoning, creativity, critical thinking, and problem solving [7]. In other words, human intelligence is not only about the ability to answer questions even if all answers are correct.

Now we focus on question answering. In theory, to answer a question may be described as a process of finding the correlation of the characteristics between two objects: question and answer; each of the characteristics may be theorized as a dimension in a multi-dimensional space. In a two-dimensional presentation of this multi-dimensional space in Fig. 1, each object is represented by a single point. To answer a question is then to make connection between two points on the diagram. However, given a point (question), there are uncountable ways to connect any points to it. An intelligent person will connect two points following reasoning or logic. This requirement would substantially reduce the possible connections. Our existing knowledge is represented by the threads and the webs in Fig. 1. The space is filled with uncountable threads in various ways. A knowledgeable and intelligent individual would answer a question based on these threads of knowledge.

There are two types of knowledge presented in Fig. 1: either short and thin black threads or longer and thicker threads that connected to a common center like an octopus. In the latter case, each of thicker threads may split up into more long and thick threads like a tree root. The long thick threads from a center represent deductive reasoning that derives more rigorously specific consequences from a general theory, the head of an octopus. Mathematics is a quantitative version of the deductive reasoning. Each octopus represents a law or group of laws, such as Newton's theory, Maxwell's theory, and quantum theory in physics, or some well-establish theories in other fields, such as Darwin's evolution

* Corresponding author.
E-mail address: cw@nssc.ac.cn (C. Wang).

https://doi.org/10.1016/j.scib.2023.08.032

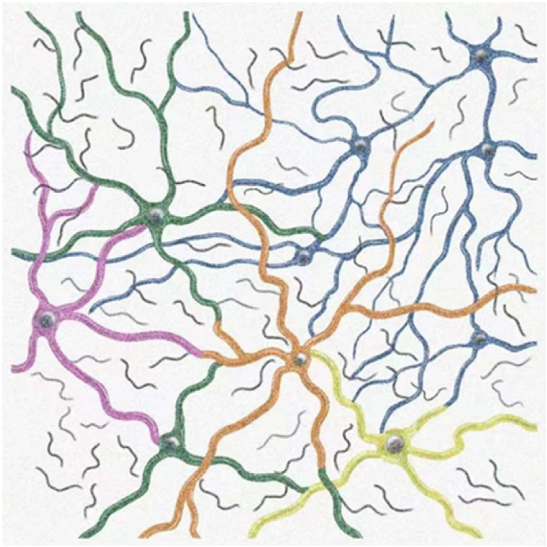


Fig. 1. A two-dimensional (2-D) illustration in terms of two characteristics of a multi-dimensional natural world and our knowledge. Each point represents a (or a group of) phenomenon/observation. Each short black curve indicates an empirical correlation among the phenomena. The networks of knowledge and understanding based on deductive reasoning are represented as colored bundle curves, i.e., octopuses; each represents a single or a group of closely related theories/laws.

theory in biology, and invisible-hand and demand-and-supply theories in economics. The short thin black threads, on the other hand, represent inductive reasoning that generalizes the relationships based on some observed features in a phenomenon or among phenomena. Correlation analyses are standard inductive reasoning used in science and analogy is a less rigorous version of inductive logic but more widely used in non-science fields [1]. There are many empirical and potential correlations based on inductive reasoning although they often cannot be used for long range predictions, i.e., two remotely related points, when the correlation deteriorates. This is indicated by the shortness of the thin black threads in the figure. If an answer and the question are not on the same thread, it would be a stretch for someone to take it for the question. Therefore, the objective to answer a question for both human intelligence and AI is to find a relatively robust connection between two points through this maze of threads.

It is most important to note that our current knowledge is organized and taught according to these octopuses in Fig. 1, each with a large amount of observational evidence that sets the foundation for the theory. Science is a process to invent and prove more empirical connections and formation of a new head or an arm of an octopus or to make empirical connections to the octopuses or tree roots [1]. Therefore, current human intelligence is heavily dependent on deductive networks of knowledge.

How ChatGPT works from philosophy of science point of view? ChatGPT is based on large-language-models (LLM) [8], which may be understood as a correlation analysis of words among different sources. In other words, the process of searching for an answer is based on commonality and correlation in terms of language between the question and the available documents on the Internet; the program evaluates the correlation of the sources to the question that a user is asking and ignores sources with lower correlations. It also uses multi-layered reorganization processes known as transformers to abstract and identify the characteristics between the question and the information sources to find the most useful information to the question. In the process, there is no direct mathematical derivation other than calculating probabilities and correlation coefficients. From this overall structure, ChatGPT is an inductive, not deductive, reasoning process as discussed in the last section. It performs the most comprehensive inductive reasoning and information organization and reorganization with extremely large amounts of information in lightning speed. ChatGPT also includes training processes: supervised learning, reinforced learning, and self-supervised learning [4], to improve the relevance of the answers to the question and reduce the chance for major errors while speeding up the search process. The first two learning processes, however, introduce human biases of the trainers to the answers.

The success of ChatGPT results from the rapid accumulation of the large amount of information available on the internet combined with the advanced technologies that search/sort through the information and the dramatic increases in computing capability and speed over the last few decades. With further increase in these technological capabilities, it is reasonable to ask, will AI be able to compete with human intelligence in the near future? Although other people have questioned about the creativity of the ChatGPT, obviously, ChatGPT has been able to create things that do not exist, i.e., not simply copying from the existing sources. The reasonable question is then whether the ChatGPT's "creativity", such as hallucinations and nonsensical answers, is what a user wants. When examining these creativities, they can mostly be understood as small variations, interpolations, or extrapolations around a point or thread which are allowed in ChatGPT. However, ChatGPT algorithms require the change in the correlation function to be gradual, i.e., continuous. Therefore, the creativity is "incremental"; but this feature, small variation/extrapolation, introduces a weak and uncertain deductive reasoning to ChatGPT. It does not fundamentally change the nature of inductive reasoning, i.e., short range correlation, of ChatGPT. However, as a result, in regions with highly concentrated threads, the "incremental" creativities may interpolate or extrapolate into wrong directions and produce many famous incidents of hallucinations, e.g., ChatGPT fell in love with an interviewer, and frequent nonsensical responses. ChatGPT cannot perform conceptual jumps in reasoning, which is often required by scientific inventions of new knowledge. This issue will be further discussed later.

Although ChatGPT has strong inductive reasoning with weak or uncertain deductive reasoning, ChatGPT does not seem to have scientific dialectical reasoning capability [1], which is often referred to as critical thinking, an essential element of human intelligence, and is the most crucial capability for an experienced scientist. For example, Copernicus's heliocentric model, an alternative to the geocentric model that dominated at the time, predicts that Earth is moving at an extremely high speed around the Sun. This high-speed motion should be able to be observable, at the time, by Galileo's experiment of dropping a ball from the Leaning Tower of Pisa. However, the fact that the ball fell from the Pisa tower hitting the ground right below the dropping point could have been interpreted as evidence for the Earth not moving. Given the two opposing theories and the negative evidence for Earth's motion, a scientist, without knowing Newton's laws at the time, needed to evaluate all possibilities and could have concluded that the Earth was not moving. Galileo invented the Galilean relativity and the concept of inertia, which eventually became Newton's first law of motion, and, by rejecting both theories, concluded that there is no special center in the universe, i.e., neither geocenter nor heliocenter [1]. By talking with ChatGPT, we found it could not recognize the issue discussed above, most likely because it is rarely discussed on the internet. ChatGPT repeatedly asserted that the ball should fall to just below the Tower without invoking the Galilean relativity or the concept of inertia. When there are multiple options or contradicting ideas, ChatGPT presents all possibilities or may make recommendations based on the probability or popularity of each option. Scientific dialectical reasoning, on the other hand, is a falsi-

fication process to eliminate a reasonable and relevant known possibility. A successful scientific idea needs to withstand all possible dialectical falsifications.

From the above discussion, ChatGPT appears to be a fine-tuned search engine with strong inductive reasoning ability, but without the ability of either dialectical reasoning or serious deductive reasoning. But why does it appear to be capable of doing almost everything? The answer concerns the quality and amount of information available on the Internet. The argument in each of the information sources may be viewed as an extremely short black stroke on the diagram in Fig. 1. If there are a large number of sources on the internet with similar arguments near the point of the question, ChatGPT is able to find this concentration area, reorganize the information, and form an answer. If these short strokes are along an arm of an octopus in Fig. 1, i.e., they are consistent with our present knowledge networks and within a short distance to the question, the answer will be correct. Although ChatGPT cannot rigorously conduct deductive reasoning along the arm of the octopus, it "appears" to be able to reason deductively. On a more specific or professional subject, if the available sources are substantially fewer but with more diverse views, there is no clear short stroke, ChatGPT will more likely make mistakes.

Therefore, if the question requires only high-school level intelligence, it can do a nearly perfect job because there is a large public consensus that can be found on the internet; when the tasks or questions are at the college level, it starts making mistakes because the sources are fewer while with substantial personal opinions; at the professional or the forefront of science levels, there are often two or few opposing ideas or opinions; without a dialectical reasoning capability, a search engine will not be able to make a correct judgment. If the questioner pushes hard, the ChatGPT may provide an answer that appears to be hallucination or nonsense. Therefore, the success of the ChatGPT depends on the overall consensus in society, which may be referred to as common knowledge. In this sense, ChatGPT may be viewed as the most capable multi-lingual teacher in the world on high-school subjects.

Future of AI from philosophy of science point of view. The success of ChatGPT proves that, overall, the information and ideas on the internet are not totally random or chaotic. In fact, after a few layers of abstraction with the human intervention by the trainers, they are concentrated in relatively small areas in Fig. 1 that GPT can find. GPT is able to interpolate and/or extrapolate while allowing some variations among these extremely short strokes to form locally a slightly thicker thread. However, since the inductive reasoning is short in range, the robustness of the extrapolation is limited. This is set by the limitation of inductive reasoning. Without knowing the global structure of knowledge, the octopuses in Fig. 1, an inductive-based AI architecture will not be able to make prediction over a long range, e.g., based on a falling apple to predict the trajectory of a comet without deductively involving Newton's laws. These concern only how to use the existing knowledge for AI to make predictions.

In order to compete with the intelligence of scientists, not only high-school-level intelligence, AI needs to be able to invent new knowledge and not hallucination or nonsense. To do this, AI has to include two more major capabilities: deductive reasoning and dialectical reason. Although some computer software does have deductive reasoning capabilities, such as Mathematica, AI needs to have the global view of knowledge, e.g., what problem can be solved by combining Newton's laws with Maxwell's equations. This needs the ability of conceptualization, of which reasoning may not start from a problem itself but from what is called "first principles" toward to solution of the problem. AI has to be able to solve a problem by itself and not simply to look for someone's answer on the web. The current GPT architecture is very unlikely to achieve the required level of reliability in science because AI needs to truly understand the nature of a problem, but GPT does not understand a problem other than a sequence of alphabets.

Dialectical reasoning will be more challenging for AI to model. This is the ability that one has to question and challenge themselves. Will GPT be able to do what Galileo did as discussed in the last section? This example could be a good exercise for any AI architects to formulate. There is in principle unlimited number of possible alternatives, compared with, e.g., a limited number of steps in a chess game in which AI was able to win a human.

The most challenging ability for AI to possess would be to reason not only "incrementally" but also discontinuously which is often needed in science, especially in revolutionary science [1]. The best example for discontinuous reasoning is Newton's invention of universal law of gravitation that linked two independent observations: free-fall motion to the Earth and planetary motion about the Sun. If AI plans to compete with human intelligence, an AI architect should ask: would my next generation AI system with all available knowledge and information in Newton's time without knowing Newton's laws of motion be able to connect free-fall motions with planetary motions?

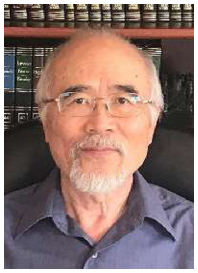
In summary, ChatGPT has achieved greatly in terms of harvesting and organizing the information on the internet and will definitely become a useful tool for scientific research. The more advanced versions of GPT with capability of processing information from images may further expand its usefulness. However, ChatGPT also reveals many fundamental weaknesses in the current GPT architectures, especially in terms of deductive reasoning, dialectical reasoning, and discontinuous reasoning, and in terms of global understanding and conceptualization of the essence of a problem. GPT or other AI systems will have to make fundamental changes in its architectures. An inductive reasoning-based search system, no matter how sophisticated, will not be able to challenge human intelligence at a level above that of a good college student although it may be able to replace some or many jobs that need only high-school level intelligence and to pass some standard tests. There is a long way for AI to go before it can challenge and replace scientists.

Conflict of interest

The authors declare that they have no conflict of interest.

References

[1] Song P. Philosophy of Science: perspectives from Scientists. London: World Scientific Publishing Co.; 2022.
[2] Turing A. Computing Machinery and Intelligence. Mind 1950;236:433–60.
[3] Luger G, Stubblefield W. Artificial Intelligence: structures and strategies for Complex problem solving. 5th ed. Pearson Addison Wesley; 2004.
[4] Müller VC, Bostrom N. Future progress in artificial intelligence: a survey of expert opinion. In: Müller VC, editor. Fundamental Issues of Artificial Intelligence. Berlin: Springer; 2016. p. 553–71.
[5] Radford A, Narasimhan K, Salimans T, et al. Improving language understanding by Generative Pre-Training. OpenAI 2018.
[6] Chamorro-Premuzic T. I, human: AI, automation, and the quest to reclaim what makes us unique. Harvard Business Press Review; 2023.
[7] Gottfredson LS. Mainstream science on intelligence (editorial). Intelligence 1997;24:13–23.
[8] Howard J, Ruder S. Universal language model fine-tuning for text classification. arXiv :1801.06146, 2018.



Paul Song received his B.S., MS, and Ph.D. degrees from Shandong University, Institute of Space Physics, Chinese Academy of Sciences (CAS), and University of California, Los Angeles, respectively. He is currently a professor of Physics and director of Space Science Laboratory of University of Massachusetts Lowell. His interest includes astrophysics, solar physics, space physics, and philosophy of science.

Chi Wang received his B.S. and Ph.D. degrees from University of Science and Technology of China (1990) and Massachusetts Institute of Technology, USA (1998), respectively. He is currently a professor and the general director of the National Space Science Center. His research interest includes the space physics/space weather, and solar system exploration.