Zehn
Vol.27 / No.107 / Autumn 2026
Home Page: zehn.iict.ac.ir
the Research Institute for Islamic Culture and Thought
Print ISSN: 1735-0743

Artificial Intelligence in the Chinese Room: An Analysis of Meaning Comprehension in Large Language Models

Hamid Mohseni[1]

1. Assistant Professor, Department of Cyberspace and Artificial Intelligence Studies, Research Institute for Islamic Culture and Thought (IICT).
E-mail: h.mohseni1297@mailfa.com

Article Info

Article type:
Research Article

Article history:
Received 2026/06/06
Received in revised form 2026/08/18
Accepted 2026/08/18
Published online 2026/09/23

Keywords:
Artificial Intelligence, Large Language Models, Meaning Comprehension, Chinese Room Argument.

ABSTRACT

The main issue of this research is to investigate the possibility of "understanding meaning" in large language models; an issue that has gained increasing importance with the expansion of the use of these models and the attribution of cognitive concepts to them. The aim of the research is to critically analyze this claim by relying on the distinction between "form" and "concept" and to analyze it with the focus on John Searle's "Chinese Room" argument. The research method is analytical-conceptual and based on the examination of philosophical texts and computational linguistics studies. The findings show that language models, despite their high ability to process and produce language, operate solely on the basis of syntactic and statistical patterns and lack access to communicative intent and reference to the outside world. Therefore, what is considered as "understanding" remains at a formal level. The result is that attributing understanding meaning to these models, in the strict philosophical sense, is not justified and the realization of meaning requires some kind of connection with the real world.

Cite this article: Mohseni, H. (2026). Artificial Intelligence in the Chinese Room: An Analysis of Meaning Comprehension in Large Language Models, *Zehn*, 27 (3), 31-54.
https://doi.org/10.22034/zehn.2026.2090523.2212

DOI: https://doi.org/10.22034/zehn.2026.2090523.2212

Extended Abstract

Introduction

The primary problem of this research is to examine the possibility of "meaning comprehension" in large language models – a problem that has gained increasing significance with the expansion of these models' applications and the attribution of cognitive concepts to them. Large language models, which refer to transformer-based models with hundreds of billions of parameters, are trained on vast amounts of textual data and attempt to probabilistically recognize and predict sequences of words that are likely to occur in human interaction. Examples such as LLaMA, GPT-3, Gopher, and PaLM demonstrate the extensive diversity of these models in terms of the number of parameters, data volume, and different applications. Although the history of chatbots dates back to the 1960s and programs like "ELIZA," the rapid development of automated software tools, along with the use of natural language processing and machine learning techniques, has enabled modern chatbots to interact and engage in virtual dialogue with human users.

However, the central problem of this article is to examine the possibility of meaning comprehension by these models, because a considerable number of researchers in the fields of artificial intelligence and natural language processing have attributed terms such as "knowledge," "knows," "believes," and similar concepts to language models. In contrast, this article, relying on the distinction between "form" (surface or structure) and "concept" (meaning), and centering on John Searle's "Chinese Room Argument," critically examines these attributions. The aim of the research is to critically analyze the claim of meaning comprehension in language models and to demonstrate the inadequacy of mere formal processing for achieving genuine understanding.

Methods

The research method in this article is **analytical-conceptual** and is based on the examination of philosophical texts and computational linguistics studies. The research begins with an explanation and examination of the distinction between form and concept, and then, centering on the Chinese Room Argument, analyzes the possibility of meaning comprehension by artificial intelligence and language models. Prior to that, the views of a considerable number of scholars who have attributed cognitive concepts to language models are briefly explained.

Results

The findings of the research indicate that language models, despite their high capability in processing and producing language, operate solely on the basis

of syntactic and statistical patterns and lack access to communicative intention and reference to the external world. Therefore, what is perceived as "understanding" remains at a formal level. According to the definition of Bender and Koller, current language models have learned only through "form" and have no way to access "meaning," because training data consist only of "form," and meaning – as a relation between form and something outside of language – requires connection to the speaker's communicative intention or conventional meaning that refers to the external world.

The Chinese Room Argument also clearly demonstrates that merely having syntax (manipulating symbols according to formal rules) is insufficient for comprehending meaning. Just as the non-Chinese-speaking person inside the room, despite providing correct and meaningful responses to Chinese questions, has no understanding of the Chinese language, language models, despite producing appropriate outputs, lack genuine understanding of the content they generate. The article emphasizes that the factor which comprehends the meaningfulness of the texts produced by the model is the human being itself; that is, it is the human agent who, by attributing communicative intention, establishes the connection between language and the external world.

Furthermore, according to Carter's analysis, although the Chinese Room Argument alone cannot completely refute computationalism, it clearly shows that mere linguistic operations, without connection to the external world, are insufficient for generating meaning, and that "embodied experience" is necessary for the development of meanings. Consequently, models that are trained solely on forms will be fragile in a sensitive training context, because they lack the ability to connect their speech to the real world.

Conclusion

The final conclusion of the research is that attributing meaning comprehension to large language models is not justified in the precise philosophical sense, and the realization of meaning requires some kind of connection with the empirical world. The findings show that language models, despite their high capability in processing and producing language, operate solely on the basis of syntactic and statistical patterns and lack access to communicative intention and reference to the external world. Therefore, what is perceived as "understanding" remains at a formal level.

Through a critical reinterpretation of Searle's argument in light of the unprecedented developments in large language models, the article demonstrates that the fundamental differences between these models and the symbolic systems Searle had in mind – including learning from data instead of following pre-written rules, the transformer architecture and attention mechanism, and the massive increase in data and parameters – not only do

not invalidate the Chinese Room Argument but elevate it to a deeper level. The extension of the Chinese Room Argument to large language models itself constitutes an innovation of this article, which has no precedent in the Islamic Republic of Iran. Although language models represent a more advanced form of artificial intelligence compared to previous ones, according to this argument, the possibility of meaning comprehension is a fundamental challenge from which artificial intelligence inside the Chinese Room cannot easily escape.

Finally, the article emphasizes that attributing epistemic concepts to large language models is, at best, metaphorical or technical in application and should not be conflated with the profound philosophical meaning of these concepts. This distinction is necessary to prevent exaggeration about the capabilities of artificial intelligence and to properly orient future research in the fields of language, cognition, and artificial intelligence. It is also suggested that future research move toward the development of models that, in addition to linguistic data, benefit from sensory interaction, multimodal perception, and connection with the real world, so that a more precise examination of the problem of "meaning" in artificial systems becomes possible.

ذهن
سال ۲۷ / شماره ۱۰۷ / پاییز ۱۴۰۵
سایت نشریه: zehn.iict.ac.ir
انتشارات پژوهشگاه فرهنگ و اندیشه اسلامی
شاپای چاپی: ۰۷۴۳-۱۷۳۵

هوش مصنوعی درون اتاق چینی؛ واکاوی درک معنا توسط مدل‌های زبانی بزرگ

حمید محسنی[۱]

۱. استادیار گروه مطالعات فضای مجازی و هوش مصنوعی پژوهشگاه فرهنگ و اندیشه اسلامی .
E-mail: h.mohseni1297@mailfa.com

اطلاعات مقاله

نوع مقاله:
مقالهٔ پژوهشی

تاریخ دریافت:
۱۴۰۵/۰۳/۱۶

تاریخ بازنگری:
۱۴۰۵/۰۵/۲۷

تاریخ پذیرش:
۱۴۰۵/۰۵/۲۷

تاریخ انتشار:
۱۴۰۵/۰۷/۰۲

واژگان کلیدی:
هوش مصنوعی، مدل‌های زبانی بزرگ، درک معنا، استدلال اتاق چینی.

چکیده

مسئله اصلی این پژوهش، بررسی امکان «درک معنا» در مدل‌های زبانی بزرگ است؛ مسئله‌ای که با گسترش کاربرد این مدل‌ها و نسبت‌دادن مفاهیم شناختی به آن‌ها اهمیت فزاینده‌ای یافته است. هدف پژوهش، تحلیل انتقادی این ادعا با تکیه بر تمایز میان «شکل» و «مفهوم» و واکاوی آن با محوریت استدلال «اتاق چینی» جان سرل است. روش تحقیق، تحلیلی-مفهومی و مبتنی بر بررسی متون فلسفی و مطالعات زبان‌شناسی محاسباتی است. یافته‌ها نشان می‌دهد که مدل‌های زبانی، علی‌رغم توانایی بالا در پردازش و تولید زبان، صرفاً براساس الگوهای نحوی و آماری عمل می‌کنند و فاقد دسترسی به قصد ارتباطی و ارجاع به جهان خارج هستند. ازاین‌رو آنچه به‌عنوان «فهم» تلقی می‌شود، در سطحی صوری باقی می‌ماند؛ نتیجه آن‌که نسبت‌دادن درک معنا به این مدل‌ها، به‌معنای دقیق فلسفی، موجه نیست و تحقق معنا مستلزم نوعی ارتباط با جهان تجربی است.

استناد: محسنی، حمید (۱۴۰۵). هوش مصنوعی درون اتاق چینی؛ واکاوی درک معنا توسط مدل‌های زبانی بزرگ، ذهن، ۲۷ (۳)، ۵۴-۳۱.
https://doi.org/10.22034/zehn.2026.2090523.2212

ناشر: پژوهشگاه فرهنگ و اندیشه اسلامی
DOI: https://doi.org/10.22034/zehn.2026.2090523.2212

مقدمه

مدل‌های زبانی بزرگ (Large Language Models) ـ با علامت اختصاری LLMs ـ همان «مدل‌های زبانی پیش‌آموزش‌دیده» ((Pre-Trained Language Models (PLM) سایز بزرگ هستند. به‌طورکلی مدل‌های زبانی بزرگ به مدل‌های زبانی ترنسفورمری اشاره دارند که حاوی صدها میلیارد (و بیشتر) پارامتر (Parameters) هستند. این مدل‌ها بر روی صدها ترابایت داده متنی آموزش داده می‌شوند (Shanahan, 2022, p). بنابر یک تعریف کاربردی، «ال‌ال‌ام‌ها»، مدل‌های زبانی عظیمی هستند که بر روی مقدار زیادی از مجموعه‌دادگان ـ بدون این‌که داده‌ها برای وظایف خاصی تنظیم شده باشند ـ آموزش دیده‌اند.

مدل‌های زبانی سعی می‌کنند به‌طور احتمالی توالی کلماتی ـ که به‌طور معمول در تعامل انسانی امکان ایجاد دارد ـ را تشخیص دهند. مدل‌ها توالی کلمات را از طریق «الگوریتم‌های مولد و تشخیصی» (generative and discriminative algorithms) و از طریق یادگیری عمیق و معماری ترنسفورمر و شبکه‌های عصبی پیش‌بینی می‌کنند (Dwivedi, et al, 2023, p 3).

مدل‌های زبانی بزرگ انواع بسیار متعدد و متنوعی دارند. برخی شاخصه‌هایی همچون تعداد پارامترها، میزان بزرگی متن آموزش‌دیده، اهداف و کاربردهای متفاوت، سیر صعودی در نسل یک مدل و شرکت‌های مختلف انتشاردهنده و عوامل دیگر باعث تنوع در این نوع مدل‌ها شده است.

به طور نمونه مدل‌های نظیر LLaMA با ۱۳ میلیارد پارامتر، GPT-3 با ۱۷۹ میلیارد پارامتر، Gopherبا ۲۸۰ میلیارد پارامتر و PaLM با ۵۴۰ میلیارد پارامتر، برخی از انواع متنوع مدل‌های زبانی بزرگ به‌حساب می‌آیند (مشاهده کنید: Touvron, et al , 2023; Brown, et al, 2020 ; Rae, et al, 2021 & Chowdhery, et al, 2023).

هرچند پیشینه «چت‌بات‌ها» (Chatbots) به دهه ۱۹۶۰ برمی‌گردد و ریشه در برنامه رایانه‌ای مانند «الیزا» (ELIZA) دارد. بااین‌وجود با پیدایش هوش مصنوعی توسعه سریع ابزارهای نرم‌افزاری خودکار نظیر چت‌بات‌ها رونق گرفت. بدین‌سان چت‌بات‌های امروزی هم از تکنیک‌های پردازش زبان طبیعی و یادگیری ماشین برای تحلیل و درک داده‌های متنی و صوت استفاده می‌کنند. این امر چت‌بات‌ها را قادر می‌سازد تا با کاربران انسانی تعامل بیشتری داشته باشند و در «گفتگوی مجازی» (virtual conversations) سریع‌تر و دقیق‌تر به انسان‌ها پاسخ دهند (Dwivedi, et al, 2023, p 22).

اما مسئله اصلی پژوهش حاضر بررسی «امکان درک معنای مدل‌های زبانی بزرگ» را تشکیل داده است؛ زیرا تعداد قابل توجهی از محققان حوزه هوش مصنوعی و پردازش زبان طبیعی واژگانی

نظیر «معرفت و دانش»، «می‌داند»، «باور»، «باور دارد» و مفاهیمی از این قبیل را به مدل‌های زبانی نسبت داده‌اند. بدین‌سان در این مقاله امکان درک و فهم معنا توسط هوش مصنوعی و مدل‌های زبانی را با محوریت «استدلال اتاق چینی» مورد و نقد و بررسی قرار می‌دهیم.

نخست فرق بین شکل (صورت یا ساختار) و مفهوم (معنا) را تبیین و بررسی کرده، سپس از منظر استدلال اتاق چینی امکان درک معنا توسط هوش مصنوعی و مدل‌ها را واکاوی می‌کنیم. البته پیش از این به‌طور مختصر اقوال تعداد قابل توجهی از اندیشمندان که مفاهیم شناختی را به مدل‌های زبانی نسبت داده‌اند، تبیین می‌شود.

۱. نسبت‌دادن مفاهیم شناختی به مدل‌ها

برخی اندیشمندان مفاهیم نظیر ادراک، فهم و درک معنا را به مدل‌های زبانی نسبت داده‌اند.

در پژوهشی دیدگاه برخی فلاسفه و دانشمندان کامپیوتر غربی چیستی دانش و امکان آن در مدل مورد بررسی قرار گرفته است (Fierro, et al, 2024, p 16102). در تحقق مزبور از حدود ۱۰۰ نفر که بیش از نصف آن‌ها فیلسوف و بقیه دانشمندان کامپیوتر بوده‌اند، نظرسنجی شده است. یکی از پرسش‌ها این بود: آیا مدل‌های زبانی بزرگ (به‌لحاظ تجربی، در عمل و اکنون) «می‌دانند»؟. در پاسخ بدین‌مسئله تفاوت معناداری بین فیلسوفان و دانشمندان کامپیوتر دیده می‌شود. بیشتر فلاسفه مخالف بوده و ۵۴ درصد نظرشان نسبت به «دانستن» فعلی مدل‌ها منفی بوده و تنها ۱۱ درصد نظرشان مثبت بوده است (بقیه هم نظری نداشتند). در مقابل ۳۱ درصد از دانشمندان کامپیوتر نظرشان منفی بوده و ۳۴ درصد دیدگاهشان این بوده که مدل‌ها «دانش» دارند.

اما سؤال دیگری که در این مقاله مطرح شده این است: آیا مدل‌های زبانی بزرگ می‌توانند به‌لحاظ نظری دانش داشته باشند؟ از منظر نظری (برخلاف عمل)، این مطلب در هر دو گروه با افزایش تأیید مواجهه شد. ۲۴ درصد از فیلسوفان پاسخ مثبت و ۳۳ درصد از آن‌ها پاسخ منفی دادند. در میان دانشمندان کامپیوتر ۵۵ درصد پاسخ مثبت داده و ۲۱ درصد نظرشان منفی بوده. این نظر مهندسان نشان‌دهنده این است که بیشتر آن‌ها معتقدند مدل‌های زبانی می‌توانند «معرفت یا دانش» داشته باشند. نتایج این نظرسنجی نشان می‌دهد که پژوهشگران از هر دو حوزه معرفت‌شناسی و علم کامپیوتر بر این باورند که مفهوم «معرفت» برای مدل‌های زبانی بزرگ موضوعی بی‌اهمیتی نیست. به‌رغم اختلاف‌نظرها، دو نکته کلیدی به‌دست می‌آید:

۱. بیشتر پژوهشگران معتقدند موجودات غیرانسانی نیز می‌توانند دانش داشته باشند.

۲. به یک معنا ال‌ال‌ام‌ها ظرفیت «دانستن» را دارند.

افزون بر این، با مطالعاتی که در حوزه ال‌ال‌ام‌ها داشتیم به دیدگاه‌های مختلفی از محققان غربی درباره امکان وجود مفاهیم شناختی و ادراک در مدل‌ها زبانی برخوردیم. در ادامه این دیدگاه‌ها را تبیین می‌کنیم.

به‌طورکلی از منظر برخی محققان انجام بسیاری وظایف پایین‌دستی مانند پاسخگویی به سؤالات و استنتاج زبان طبیعی براساس «درک» رابطه بین دو جمله استوار بوده که این به‌طور مستقیم توسط مدل‌سازی زبان به تصویر کشیده نمی‌شود. به همین خاطر برخی محققان سعی کرده‌اند این مشکل را در مدل‌های اولیه نظیر BERT حل کنند. آن‌ها برای رفع این مشکل این ایده را مطرح کرده‌اند که مدل را به‌گونه‌ای آموزش بدهند تا روابط جمله را «بفهمد» (understands) (Devlin, et al, 2019, p 4174). همچنین در برخی تحقیقات با طرح اصطلاح «درک» (Comprehension) سعی شده با به‌کارگیری روش‌های فرایند «درک ماشین گفتگویی» (Conversational machine comprehension (CMC)) در این مدل تقویت شود (Ohsugi, et al, 2019). یا در برخی تحقیقات بحث «دانش واقعی» (factual knowledge) مطرح شده است که مدل‌ها توانایی فوق‌العاده‌ای برای یادآوری دانش واقعی بدون هیچ‌گونه تنظیم دقیقی دارند. این ظرفیت مدل‌ها را به‌عنوان سیستم‌های پاسخ به سؤالات «دامنه باز بدون نظارت» (unsupervised open-domain.) نشان می‌دهد. برخی تحقیقات به‌دنبال این هستند مدل‌ها را به‌گونه‌ای تقویت کنند که همانند انسان یاد بگیرند، بازنمایی ادراکات داشته و بتوانند استدلال کرده و برنامه‌ریزی داشته باشند (LeCun, 2022).

همچنین در رسانه‌ها، روزنامه‌ها و مطبوعات معروف مشاهده می‌شود فهم و درک را به مدل‌های زبانی نسبت داده‌اند. به‌طور نمونه گفته ادعا شده مدل «برت» به‌گونه‌ای آموزش‌دیده تا بتواند جستجوها را «بفهمد»؛ بدین‌سان جستجوها آسان شده و نتایج بهتری به‌بار داشته است؛[1] درحالی‌که برخی در استفاده از واژگان معرفتی درباره مدل‌های زبانی در رسانه‌های عمومی احتیاط می‌کنند. به‌طور نمونه در روزنامه **نیویورک تایمز** آمده هنوز راه طولانی هست تا مدل‌ها بتوانند «فهم واقعی» داشته باشند؛ اگرچه مدل زبانی نظیر برت توانسته آزمون «فهم مشترک» (common-sense) آزمایشگاه را پشت‌سر بگذارد، ولی هنوز با نمونه مصنوعی «فهم مشترک» فاصله زیادی دارد.[2] از منظر جان مک‌کارتی وقتی می‌توان گفت یک برنامه «فهم عمومی» دارد که به‌طور خودکار با توجه به آنچه به او گفته شده و آنچه از قبل می‌دانسته، بتواند استنباط کند (McCarthy, 1959 , p 4).

1. https://blog.google/products/search/search-language-understanding-bert/
2. https://www.nytimes.com/2018/11/18/technology/artificial-intelligence-language.html

از منظر تحقیقات دیگری، درک ناقص جهان فراتر از متن، درک ناکافی رفتار انسانی و درک ناکافی از تفاوت‌های فرهنگی، از چالش‌های اساسی هستند که در استقرار ایمن، مسئولانه و کارآمد ال‌ال‌ام‌ها تأثیرگذار خواهند بود (Hadi, et al, 2023, p 22).

به باور برخی محققان، مدل‌های زبانی نظیر GPT-3 بخشی از فناوری خارق‌العاده بوده و به‌عنوان یک ماشین‌تحریری قدیمی به‌حساب می‌آیند، اما ماشینی «باهوش، آگاه، زیرک، بیدار، ادراکی و بصیر، حساس و معقول» (intelligent, conscious, smart, aware, perceptive, insightful, sensitive and sensible) هستند؛ هرچند این واژگان مستلزم وجود مفاهیم شناختی، ادراکی و احساسی در مدل‌های زبانی است. با این وجود این محققان تذکر می‌دهند هوش مصنوعی شبیه آنچه هالیوود نمایش می‌دهد تنها در فیلم‌ها پیدا می‌شود (Floridi & Chiratti, 2020, p 690) و هنوز مدل GPT-3 آن‌قدر قوی نبوده و خروجی‌ای همانند انسان تولید نکرده و فاقد «فهم مشترک» است؛ به‌نظر می‌رسد کارهایی که این نوع مدل انجام می‌دهد شبیه کارهای برش و چسباندن باشد تا این‌که آثار اصیل تولید کند (Will Douglas Heaven, 2020).

همچنین از منظر بندر و کولر اگر مراد از اصطلاحاتی نظیر فهم، درک و یا یادآوری دانش واقعی، مشابهت با انسان باشد، این ادعاها به‌شدت مبالغه‌آمیزند، اما اگر منظور از این واژگان اصطلاحات فنی در نظر گرفته شده باشند، می‌بایست این اصطلاحات به‌صراحت تعریف شوند (Bender and Koller, 2020, p 5186).

تا بدینجا دیدگاه محققان و اندیشمندان مختلف در حوزه درک معنا توسط مدل‌های زبانی تبیین شد.

۲. فرق شکل و مفهوم

امیلی ام. بندر ـ استاد زبان‌شناسی و عضو هیئت علمی دانشگاه واشنگتن در گروه زبان‌شناسی و استاد کمکی در گروه علوم و مهندسی کامپیوتر[1] ـ و الکساندر کولر ــ پروفسور زبان‌شناسی محاسباتی در بخش علوم زبان و فناوری در دانشگاه زارلند ـ تفاوت بین «شکل» (فرم (Form)) و «مفهوم» در سیستم‌ها را موردتوجه قرار داده‌اند (Bender and Koller, 2020). البته این دو محقق بیشتر از اصطلاح مفهوم «Meaning» استفاده کرده و به‌ندرت اصطلاح معنا «Semantics» (سمنتیک) را در مقاله خودشان به‌کار برده‌اند و برخی جاها هر دو واژه را پشت‌سر

1. https://linguistics.washington.edu/people/emily-m-bender.

هم آورده که به‌نظر می‌رسد در نگرش آن‌ها این دو اصطلاح در مورد سیستم‌ها بـه یـک معنـا گرفتـه شده است، اما در فلسفه تحلیلی و زبان‌شناسی دیدگاه‌های مختلفـی دربـاره معنـا و مفهـوم وجـود دارد. ما در این مقاله وارد این اختلاف دیدگاه نشده و مفهوم و معنا در حوزه مدل‌های زبانی را یکـی گرفته و با محوریت نگاه این دو محقق زبان‌شناس و دانشمند علوم کامپیوتری آن مفاهیم را تبیـین و بررسی می‌کنیم.

به‌طورکلی مراد محققان مذکور از «معنا یا مفهوم»، «معنـای زبـانی» بـوده کـه بـه رابطـه بـین «شکل زبان» و «قصد ارتبـاطی» (communicative intent) اشـاره دارد. آن‌هـا مـدعی هـستند مدل‌های زبانی فعلی به‌دلیل این‌که فقط از طریق «شکل» یاد گرفته‌اند، هیچ راهی برای دست‌یـابی به «معنا» ندارند. آن‌ها تذکر می‌دهند درک واضح از تمایز بین «شـکل» و «معنـا» بـه جهـت‌دهـی رشته زبان طبیعی به‌سمت دانش بهتر کمک خواهد کـرد. هرچنـد در تحقیقـات دانـشگاهی هنـوز مشخص نیست کـه آیـا همـه نویـسندگان بـه تمـایز بـین شـکل و مفهـوم ملتفـت هـستند یـا نـه؟ (Bender and Koller, 2020, p 5185-86)

ازاین‌رو این محققان زبان‌شناس با بررسی مقـالات دربـاره «بـرت لـوژی» (BERTology) و نحوه عملکرد این مدل زبانی در پیشین‌بینی کلمات و استنتاج، بدین‌نتیجه می‌رسند مدل‌های زبانی بزرگ می‌توانند ساختار شکلی زبان ـ به‌طور نمونه توافق و ساختار وابستگی ـ را بیاموزند، اما درباره توانایی این سیستم‌ها در «استدلال‌کردن» (Reason)، به‌گونه‌ای توهم شده است. به‌بیان‌دیگر ایـن نوع استدلال‌کردن مدل‌ها، براساس «شکل» است نه مبتنی بـر «مفهـوم». ازایـن‌رو بـه بـاور آن‌هـا، مدلی که براساس «شکل» (فرم) آموزش دیده است علی‌الاصول نمی‌تواند «مفهوم» را یـاد بگیـرد؟ (Bender and Koller, 2020, p 5185).

البته بندر و کولر به موفقیت‌های مدل‌های زبانی عصبی بزرگ در وظایف پردازش زبان طبیعـی اعتراف کرده‌اند. هرچند برخی دست‌یابی به‌معنا توسط مدل‌ها را توضـیح مـی‌دهنـد، ولـی ایـن دو تصریح می‌کنند این امر منجر به «درک زبان» (understanding language) و یا دسـت‌یـابی بـه «معنا» توسط مدل‌ها نمی‌شود. بندر و کولر در این پژوهش اثبات می‌کننـد سیـستمی کـه براسـاس «شکل» آموزش‌دیده است، هیچ راهی برای یادگیری معنا ندارد.

همچنین از منظر فیلسوف ذهن جان سرل سیستم‌های مبتنی بر «شکل» قادر به درک «مفهـوم و معنا» نیـستند (Searle, 1980, p 11). بـه‌بیـان‌دیگـر سـرل تـلاش مـی‌کنـد ثابـت کنـد «نحـو» (Syntax) بـــه تنهـــایی بـــرای دســـت‌یـــابی بـــه «معنـــا» (Semantic) کفایـــت نمـــی‌کنـــد (Carter, 2007, p175).

همچنـین از منظـر برخـی کتـب زبـان‌شناسـی زبـان بـه‌عنـوان سیـستم‌هـای نـشانه اسـت

(Saussure, 1959)؛ یعنی زبان از جفت‌شدن شکل و معنا به‌دست می‌آید، اما داده‌های آموزش برای مدل‌های زبانی فقط «فرم» هستند؛ بنابراین ادعاهای مربوط به توانایی‌های مدل در درک معنا و مفهوم می‌بایست با دقت ملاحظه شوند (Bender, et al, 2021, p 615).

۲-۱. تعریف شکل و مفهوم در مدل‌ها

دو محقق زبان‌شناس بندر و کولر، شکل (فرم) و معنا در مدل‌ها را این‌گونه تعریف کرده‌اند:

شکل: منظور از شکل یا ساختار زبان، هرگونه ظهور قابل مشاهده از زبان است؛ همانند علایم صفحه، پیکسل‌ها و یا بات‌ها در بازنمایی دیجیتال متن، یا حرکت اندام‌های گفتاری، در زمره شکل یا ساختار زبان قرار می‌گیرند.

معنا: معنا به‌عنوان رابطه‌ای بین «شکل» و چیزی خارج از زبان در نظر گرفته می‌شود (Bender and Koller, 2020, p 5186-5187). در ادامه با تبیین دو مفهوم از معنا، این مطلب بیشتر تبیین می‌شود:

۲-۱-۱. مفهوم و قصد ارتباطی

هنگامی که انسان از زبان استفاده می‌کند، از این کار هدفی دارد. انسان برای صرف لذت‌بردن اندام گفتاری خودش را حرکت نمی‌دهد، بلکه برای دست‌یابی به برخی مقاصد ارتباطی صحبت می‌کند. انسان انواع مختلف مقاصد ارتباطی دارد. یک نوع قصد ارتباطی این است که برای انتقال اطلاعات برای شخص دیگر از زبان استفاده می‌کند، یا این‌که وقتی انسان قصد درخواست از دیگری دارد، این کار را انجام می‌دهد و یا این‌که زمانی تنها قصد معاشرت با دیگران دارد. در این صورت معنا، به ارتباط بین دوگانه «عبارت زبان طبیعی» و «قصد ارتباطی» (communicative intent) تعریف می‌شود. بنا بر این تعریف می‌توان منظور از مقوله «فهم» را نیز مشخص کرد. فهم به‌معنایی بازیابی قصد ارتباطی از عبارات زبانی طبیعی است (Bender and Koller, 2020, p 5187). به‌بیان‌دیگر فرایند فهم، زمانی محقق می‌شود که فرد بتواند قصد گوینده ـ از عبارات زبانی که به‌کار برده ـ را بازیابی کند.

مقاصد ارتباطی در مورد چیزی خارج از زبان هستند. به‌طور مثال وقتی گفته می‌شود «پنجره را باز کن»، یا «حافظ در چه قرنی متولد شده است؟» مقاصد ارتباطی در این موارد ریشه در دنیای واقعی دارد که گوینده و شنونده با هم در آن زندگی می‌کنند. همچنین اهداف ارتباطی می‌توانند در جهان‌های انتزاعی باشند؛ به‌طور مثال در حساب‌های بانکی، فایل سیستم‌های کامپیوتری یا در یک دنیای کاملاً فرضی در ذهن گوینده باشند (Bender and Koller, 2020, p 5187).

زبان‌شناسان «قصد ارتباطی» را از «معنای قراردادی (یا ثابت)» (conventional (or standing) meaning) متمایز می‌دانند. معنای قراردادی یا متعارف یک عبارت (کلمه و جمله) چیزی است که در تمام زمینه‌هایی ممکن برای استفاده، ثابت می‌ماند. معنای قراردادی یک چیز انتزاعی است که ظرفیت ارتباطی یک «شکل» را نمایندگی می‌کند؛ با توجه به سیستم زبانی که از آن گرفته شده است. هر سیستم زبانی (مانند زبان انگلیسی) رابطه‌ای را ارائه می‌دهد که شامل دوگانه عبارت و معنای قراردادی است. حوزه معناشناسی زبانی، نظریه‌های رقابتی بسیاری درباره این‌که «معنای قراردادی» چگونه به‌نظر می‌رسد ارائه می‌دهد، اما از منظر برخی محققان نیازی به این نظریه‌ها مختلف در اینجا نیست. تنها این نکته مفروض گرفته می‌شود که «معنای قراردادی» می‌بایست «تفسیرهایی» داشته باشد. به‌بیان‌دیگر هر کلمه یا عبارت می‌بایست روش‌هایی برای درک و تفسیر آن‌ها وجود داشته باشد تا این‌که بتوان معنای یک جمله و عبارت را فهم کرد و ارتباط آن را با واقعیت‌های جهان درک کرد. ازاین‌رو همانند رابطه معناشناسی «قصد ارتباطی»، رابطه «معنای قراردادی»، زبان را به اشیاء خارجی از آن متصل می‌کند (Bender and Koller, 2020, p 5187).

۲-۲-۲. معنا و هوش

معنا و درک همواره به‌عنوان کلیدهای هوش در نظر گرفته می‌شده‌اند. تورینگ (۱۹۵۰) استدلال کرد اگر داور انسانی پس از یک گفتگوی نوشتاری دلخواه نتواند بین ماشین و هم‌صحبت انسانی تشخیص دهد که کدام ماشین است و کدام انسان؛ در این صورت می‌توان گفت ماشین دارای تفکر است و «می‌اندیشد» (think). به همین خاطر انسان‌ها به‌راحتی تمایل دارند معنا و حتی هوش را به عوامل مصنوعی نسبت دهند؛ چنانچه از طرز رفتار مردم با مدل «الیزا» مشخص است (Bender and Koller, 2020, p 5187-8).

بدین‌سان هوش مصنوعی و مدل‌های زبانی به‌دلیل این‌که رفتار هوشمندانه از خود نشان می‌دهند می‌تواند مفاهیم شناختی مانند تفکر و ادراک معنا را به آن‌ها نسبت داد.

۳. استدلال بر عدم درک معنا توسط مدل‌ها

در این بخش این مسئله مهم بررسی می‌شود: آیا محاسبات رایانه‌ای می‌توانند جنبه مهم از عالم ذهنی یعنی معناداری حالات ذهنی انسان را توضیح دهند؟ زیرا ویژگی اصلی حالات ذهنی این است که «محتوای معنایی» (semantic content) دارند. به‌بیان‌دیگر حالات ذهنی، «حالات

مفهوم‌دار» (meaningful states) هستند و هر نظریه معقولی از ذهن می‌بایست بتواند محتوای معنایی حالات ذهنی را توضیح دهد، اما محاسبات یک فرایند کاملاً «نحوی (یا نشانه‌شناختی)» (Syntactic) است؛ عملیات سیستم‌های شکلی (یا صوری) به‌طور نحوی و به‌عنوان دست‌کاری نمادها مشخص شده‌اند. بدین‌سان مسئله اصلی که می‌بایست بررسی شود: آیا معنا از طریق عملیات نحوی صرف توجیه می‌شود. جان سرل تلاش کرد با آزمون فکری ابداعی خود اثبات کند «نحو» (syntax) به تنهایی برای «معنا» (Semantic) کفایت نمی‌کند (Carter, 2007, p175).

ازاین‌رو در ادامه به‌طور مفصل به تبیین و بررسی استدلال جان سرل می‌پردازیم.

۳-۱. استدلال اتاق چینی

جان سرل ـ فیلسوف ذهن ـ آزمایش فکری را در سال ۱۹۸۰ در مقاله معروف «ذهن‌ها، مغزها و برنامه‌ها» ترسیم کرد که به استدلال «اتاق چینی» (Chinese Room) معروف شد. هرچند این استدلال و نقد آن در فلسفه ذهن و بررسی هوش مصنوعی قوی بسیار مورد توجه قرار گرفته است، اما نقد امکان درک معنا توسط مدل‌های زبانی از منظر این استدلال کاری بدیع به‌حساب می‌آید.

جان سرل این استدلال را در نقد تحقق هوش مصنوعی قوی مطرح کرد. وی هوش مصنوعی را به «ضعیف» و «قوی» تقسیم کرد و با تبیین ویژگی‌هایی برای این دو، آن‌ها را از هم متمایز کرد. از منظر سرل «هوش مصنوعی ضعیف» ("weak" AI) یا «محتاط» (Cautious) ابزاری بسیار قدرتمندی برای مطالعه ذهن و قوای شناختی است که انسان را قادر می‌سازد تا فرضیه‌ها را به‌طور دقیق‌تر تدوین و آزمایش کند. همچنین این نوع هوش ابزارهایی هستند که می‌توان با آن‌ها تفسیرهای روان‌شناختی را آزمایش کرد. سرل در مقاله معروفش اصلاً منتقد و منکر وجود این نوع هوش مصنوعی نیست.

اما بنابر دیدگاه سرل، «هوش مصنوعی قوی» ("strong" AI)، صرفاً کامپیوتر و ابزاری برای مطالعه ذهن نبوده، بلکه اگر رایانه‌ای به‌طور مناسب برنامه‌ریزی بشود، واقعاً «ذهن» (Mind) خواهد داشت. در این صورت می‌توان گفت رایانه «می‌فهمد» و به‌معنای واقعی کلمه «حالات شناختی» (cognitive states) نیز دارد؛ بنابراین سرل سعی دارد این نوع هوش مصنوعی قوی را نقد کند (Searle, 1980, p2).

وقتی درباره داستانی از سیستم‌های هوشمند پرسیده می‌شود و آن‌ها پاسخ می‌دهند، در این صورت از منظر طرفداران هوش مصنوعی قوی دو مطلب ذیل صحیح است:

۱. ماشین به‌طور واقعی داستان را «می‌فهمد» و برای آن پاسخی را فراهم می‌کند.

۲. کاری که ماشین و برنامه آن انجام می‌دهد، توانایی انسان برای درک داستان و پاسخ به سؤالات مربوط به آن را توضیح می‌دهد.

سرل برای بررسی دو ادعای بالا، آزمایش خود را مطرح کرد. بررسی مطلب دوم فعلاً در راستای بحث ما نیست، اما مطلب اول و بحث فهم واقعی سیستم به‌طور مستقیم با بحث حاضر ارتباط دارد و به‌نظر می‌رسد محور و تمرکز اصلی استدلال اتاق چینی نیز نقد همین مطلب است.

جان سرل فرض می‌کند خودش در یک اتاق حبس شده و دسته بزرگی از برگه نوشته چینی به او داده می‌شود، درحالی‌که سرل هیچ چیزی درباره زبان چینی نمی‌داند ـ چه نوشتاری و چه گفتاری و چه به‌لحاظ تشخیص علایم چینی از زبان‌های مشابه آن ـ؛ برای سرل زبان چینی یک‌رشته کلمات با پیچ‌وتاب‌های بی‌معناست(Searle, 1980, p 2).

ادامه استدلال اتاق چینی جان سرل را با تقریر روشن‌تری از مت کارتر توضیح می‌دهیم. تصور کنید از شما خواسته شده تا برای چند ساعتی در یک آزمایشی شرکت کرده و یک وظیفه خاص را انجام دهید. شما به اتاق بزرگی معرفی می‌شوید که شامل هزاران قفسه کتاب شماره‌گذاری شده است. در وسط اتاق میزی قرار دارد و کتابی بر روی آن است. شما آن کتاب را ورق می‌زنید و در آن فقط قوانین بازنویسی برای نمادها را مشاهده می‌کنید؛ درحالی‌که شما هرگز آن‌ها را ندیده بودید. در ضمن کنار هر قانون بازنویسی یک یادداشت عددی نیز وجود دارد.

به شما گفته می‌شود در این اتاق تنها خواهید ماند. بعد از آن یک‌تکه کاغذ با رشته نمادهای از شکاف درب وارد می‌شود. وظیفه شما این است که قاعده‌ای را در کتاب پیدا کنید که سمت ورودی‌اش دقیقاً همان رشته نمادها باشد و رشته خروجی آن قانون را بر روی طرف دیگر تکه‌کاغذ رونوشت کرده و آن را از طریق شکاف در به بیرون بگردانید. سپس شما باید کتابی را در قفسه‌ها پیدا کنید که شماره‌اش با یادداشت عددی کنار قاعده‌ای که تازه دنبال کرده‌اید مطابقت دارد و کتاب روی میز را با این کتاب جدید جایگزین کنید. دوباره یادداشت از شکاف درب وارد می‌شود. شما به کتاب روی میز نگاه می‌کنید تا این رشته را در سمت ورودی قاعده پیدا کرده و رشته خروجی آن را رونوشت کنید. سپس کتاب روی میز را با کتابی از قفسه‌هایی که شماره‌اش با یادداشت کنار قاعده داده شده تطابق دارد، جایگزین می‌کنید. تکه‌کاغذ دیگری با رشته جدیدی از نمادها از طریق شکاف درب به شما داده می‌شود و شما این رویه را تکرار می‌کنید.

پس از گذشت چند ساعت از انجام این کار، به شما گفته می‌شود که رشته‌های نمادها در واقع جملاتی به زبان چینی بوده‌اند. کتاب‌های موجود در اتاق حاوی تمام مکالمات ممکن شما به زبانی چینی در چند ساعت حضور شما در آنجاست. هر کتاب نمایانگر یک وضعیت مکالمه و پاسخ‌های معقولی برای ورودی‌های ممکن را فراهم می‌آورد. به‌نظر می‌رسد شما چند ساعت با یک

چینی‌زبان بومی گفتگو کرده‌اید و تنها با پیروی از قوانین بازنویسی موجود در کتاب‌ها، آزمون تورینگ چینی را پاس کرده‌اید، اما واضح است که شما در نهایت کار، زبان چینی نمی‌دانستید. از طرف دیگر ممکن است پاسخ‌هایی که می‌دادید با باورهای شما مطابقت نداشته باشد، بلکه با پاسخ‌های غیرموجهی که در کتاب‌ها رمزگذاری شده بودند انطباق داشته‌اند. به‌عنوان مثال شاید یکی از سؤالات این بوده: «آیا بستی توت‌فرنگی دوست دارید؟» و شما پاسخ داده‌اید: «بله، خوشمزه است»؛ درحالی‌که شما حتی تحمل دیدن بستی توت‌فرنگی را هم ندارید. یا این‌که یکی از این سؤالات این بوده «آیا گرسنه‌ای؟» شما پاسخ داده‌اید: «خیر در حال حاضر خوبم، متشکرم». با وجود این‌که شما بسیار گرسنه بوده‌اید و فکر می‌کردید کی کارتان تمام می‌شود تا بتوانید غذا بخورید.

اما صرف‌نظر از این، شما در ظاهر یک گفتگوی معقول و درستی با یک چینی‌زبان انجام داده‌اید، درحالی‌که توانایی شما برای گفتگو به زبان چینی فراتر از اتاق چینی نمی‌رود. اگر یک زبان‌دان چینی پس از بیرون آمدن از اتاق، یک سؤال کتبی به زبان چینی به شما بدهد و رشته‌ای به‌دست شما برساند، رشته نمادها برای شما بی‌معنا خواهند بود؛ زیرا تنها از طریق مراجعه به وضعیت‌های مکالمه رمزگذاری شده در کتاب‌های اتاق چینی است که شما می‌توانید پاسخ صحیحی بدهید و به‌ظاهر درک و فهم زبان چینی را از خود به نمایش بگذارید و به اصطلاح آزمون تورینگ مربوطه را پاس کنید.

این وضعیتی که در آزمایش فکری توصیف شده، همانند وضعیتی است که در فرایند آن یک سیستم صوری نمادها را طبق قوانین صوری بازنویسی می‌کند تا به‌خوبی آزمون تورینگ را پشت‌سر بگذارد. بااین‌حال شهودی که از این آزمایش فکری نشئت می‌گیرد این است که هرچند به‌ظاهر درکی توسط سیستم نمایان شده است ـ به‌اندازه‌ای که یک انسان را قانع کند ـ، اما فرایند ادراک واقعی در عملکرد سیستم وجود ندارد. آن‌ها به‌خودی خود هیچ «معنایی» (mean) ندارند. به‌عبارت‌دیگر هرچند عملیات «نحوی یا صوری» (Syntactic) اتاق چینی، آزمون تورینگ را پشت‌سر می‌گذارند، اما فاقد «معنا» (Semantics) هستند (Carter, 2007, p 175-177).

ازاین‌رو جان سرل استعاره‌ای از «سیستم» را توسعه می‌دهد که در آن شخصی وجود دارد چینی صحبت نمی‌کند، اما با مراجعه به کتابخانه‌ای از کتاب‌های چینی، طبق قوانین از پیش تعریف شده، سؤالات چینی را پاسخ می‌دهد؛ اگرچه در واقع هیچ درک واقعی در هیچ کجای سیستم اتفاق نیفتاده، ولی از بیرون به‌نظر می‌رسد سیستم معنای کلمات چینی را «می‌فهمد» (Bender and Koller, 2020, p 5188).

از منظر سرل ماشین تنها «نمادهای صوری» (formal symbols) را جابه‌جا می‌کند بدون

این‌که درکی از معنای آن‌ها داشته باشد. همانند فرد غیرچینی‌زبان داخل اتاق بدون این‌که فهمی از زبان چینی داشته باشد، تنها نمادها را جابه‌جا می‌کرد، اما ناظر خارجی فکر می‌کرده این فرد زبان چینی را می‌فهمد که به سؤالات او پاسخ‌های صحیح و معناداری می‌دهد.

خلاصه این‌که از منظر جان سرل تنها از طریق داشتن «نحو (یا نشانه‌شناسی)» (Syntactic) نمی‌توان داشتن درک «معنا» را اثبات کرد. به‌بیان‌دیگر سرل نکته اصلی برهان خوش را این می‌داند «مدل صوری» (formal model) به تنهایی و به‌خودی‌خود نمی‌تواند «دربارگی» (Intentionality) و حالات ذهنی داشته باشد. به تعبیر دیگر از برنامه‌های صوری «ویژگی‌های علّی» (causal properties) به‌وجود نمی‌آیند (Searle, 1980, p 11). البته جان سرل امکان تفکرکردن را به‌طورکلی از هوش مصنوعی و رایانه‌های هوشمند سلب نمی‌کند.

۳-۲. بررسی استدلال اتاق چینی

آزمایش فکری اتاق چینی در فلسفه ذهن و در نقد و اثبات برخی دیدگاه‌ها بسیار مورد استفاده قرار گرفته است. دیدگاه‌های مختلفی درباره ماهیت ذهن انسان در فلسفه ذهن وجود دارد؛ یکی از این رویکردها، رویکرد «محاسبه‌گرایی» (computationalism) در مورد ذهن و حالات ذهنی انسان است. استدلال اتاق چینی ممکن است دیدگاه «محاسبه‌گرایی» در مورد ذهن را به چالش بکشد. بدین‌سان مت کارتر ظرفیت این استدلال در نقد دیدگاه «محاسبه‌گرایی» را بررسی کرده است؛ به‌دلیل تناسب بحث مذکور با بحث درک معنا توسط مدل‌ها در ادامه به تبیین این مطلب می‌پردازیم.

کارتر اشکال مزبور بر دیدگاه محاسبه‌گرایی را در قالب استدلال زیر تقریر می‌کند:

۱. داشتن معنا، شرط لازم برای داشتن ذهن است.

۲. عملیات نحویِ سیستم‌های صوری، برای داشتن معناشناسی کافی نیستند.

در نتیجه: عملکرد سیستم‌های صوری برای داشتن ذهن کافی نیست.

بنابراین: دیدگاه محاسبه‌گرایی نادرست است.

مقدمه اول مورد اختلاف نیست. یعنی روشن است که حالات ذهنی محتوای معنایی دارند.

اما مقدمه دوم مبتنی بر استدلال اتاق چینی سرل است. این مقدمه را می‌توان مورد بحث و تردید قرار داد. آیا مقدمه دوم آن‌قدر کلیت دارد که همه تفسیرها از سیستم صوری را دربرگیرد؟ از این مقدمه می‌توان دو تفسیر ارائه داد:

۱. تفسیر ضعیف: دست‌کم یک سیستم صوری وجود دارد که به‌طورکلی عملیات آن برای داشتن معنا کافی نیست.

این تفسیر خیلی پیش‌پاافتاده است و مسلم است بسیاری از سیستم‌های صوری وجود دارند که عملیات آن‌ها داشتن معنا و ذهن را پشتیبانی نمی‌کنند. آزمایش فکری اتاق چینی دست‌کم اثبات می‌کند به‌طور شهودی این‌گونه سیستم‌ها به‌طور کامل فاقد درک هستند، اما آنچه جالب است با عدم درک معنا، این‌گونه سیستم‌ها هنوز می‌توانند آزمون تورینگ را پاس کنند. در این صورت می‌توان گفت آزمون تورینگ، معیار قابل اعتمادی برای داشتن ذهن نیست.

۲. تفسیر قوی: هیچ سیستم صوری وجود ندارد که عملیات آن برای داشتن معنا کافی باشد.

اما استدلال اتاق چینی قادر نیست تفسیر قوی ـ این‌که «هیچ سیستمی وجود ندارد که برای داشتن معنا کافی باشد» ـ را بتواند اثبات کند. در نتیجه از منظر مت کارتر با نقض و تردید در مقدمه دوم استدلال، نمی‌توان نادرستی محاسبه‌گرایی را اثبات کرد.

به‌بیان‌دیگر آزمایش فکری اتاق چینی به‌لحاظ منطقی امکان‌پذیر است، اما یک آزمایش فکری به‌لحاظ فیزیکی، کار فلسفی خاصی انجام نمی‌دهد؛ زیرا اندیشمندان از آزمایش فکری برای آزمودن ادعاهای روابط منطقی استفاده می‌کنند. به‌طور مثال فیزیکالیسم ادعا می‌کند یک توصیف فیزیکی کامل، به‌عنوان یک توصیف کامل از ذهن، کافی است. این کفایت یک ادعای منطقی است؛ بنابراین وجود دانشمندی مانند «مری»[1] نیز استدلال منطقی است که در برابر این کفایت منطقی استدلال می‌کند. به‌عبارت‌دیگر استدلال منطقی تنها توانایی اثبات ادعای منطقی را دارد.

از منظر مت کارتر اگر آزمون تورینگ یک آزمون منطقی بود در این صورت آزمون فکری اتاق چینی می‌توانست به‌طور واقعی برضد آن استدلال کند، اما آزمون تورینگ یک آزمون تجربی است. ازاین‌رو آزمایش فکری اتاق چینی هیچ انتقاد اثربخشی بر آزمون تورینگ ارائه نمی‌دهد؛ پس آزمایش فکری اتاق چینی به‌راحتی نمی‌تواند نادرستی محاسبه‌گرایی را نشان دهد.

اما هنوز از منظر کارتر آزمون فکری اتاق چینی اهمیت دارد. این آزمون نشان می‌دهد که هر عملیات زبان‌شناختی، بدون ارتباط با دنیای خارجی برای تولید معنا کافی نیست. بر فرض فردی می‌تواند سال‌ها در اتاق چینی حضور داشته و به همه کتاب‌ها دسترسی داشته باشد، اما باز هم نتواند معنای نمادهایی را که پردازش می‌کند، درک کند. دلیلش آن است که این عملیات فاقد ارتباط مناسب با دنیای خارجی است.

---

۱. فرد دانشمندی که در اتاق سیاه و سفیدی قرار دارد و به‌خوبی رنگ‌ها را می‌شناسد، اما تا به حال خود رنگ‌ها را ندیده است. یک روز، وقتی از اتاق خارج می‌شود و به‌طور نمونه رنگ قرمز گوجه‌فرنگی را مشاهده می‌کند، با دیدن قرمزی گوجه‌فرنگی دانش جدیدی را کشف می‌کند. این بیانگر آن است که توصیف فیزیکی صرف از یک پدیده ممکن است کافی نباشد؛ در نتیجه پدیده‌های همانند کوالیا و حالات ذهنی را نمی‌توان تنها با توصیف فیزیکی و به‌طور کامل توضیح داد.

به‌عبارت‌دیگر برای درک زبان چینی باید فهمید عناصر زبان ـ چه نوشتاری و چه گفتاری ــ چگونه به چیزهای خارج از سیستم زبان مربوط می‌شوند. زبان‌ها سیستم‌هایی هستند که معانی را رمزگذاری کرده و ارتباط برقرار می‌کنند. بااین‌حال معانی به‌طور کامل از طریق مکانیزم‌ها و ورودی‌های که کاملاً درون ساختار زبان قرار دارند، تولید نمی‌شوند. هرچند به‌طورقطع مکانیزم‌های زبانی در انتقال معنا به موجودات زبانی نقش دارند، اما باید ارتباط مناسبی با دنیای خارجی نیز وجود داشته باشد.

بنابراین از منظر مت کارتر نکته اساسی که از آزمایش اتاق چینی به‌دست می‌آید این است که «تجربه بدنمند» (embodied experience) برای توسعه معانی ضروری است؛ همان‌طور که حالات ذهنی انسان وقتی مفهوم دارند که انسان یک تجربه قبلی در جهان داشته باشد. این فرایند در انسان توسط دستگاه‌های حسی حاصل می‌شود. به‌عبارت‌دیگر معنا به تنهایی توسعه نمی‌یابد و این توسعه مشروط به تجربه‌ای در ارتباط با جهان تجربی است. کارتر معتقد است ضرورت تجربه بدنمند برای توسعه معنا، به‌خودی خود علیه محاسبه‌گرایی بحث نمی‌کند. این امر تنها محاسبه‌گرایی را ملزم می‌کند تا توضیحاتی درباره مکانیزم اعطاکننده معنا، ارائه بدهد و تبدیل داده‌های حسی به انواع مختلفی از بازنمایی‌های ذهنی را توصیف کند. به‌عبارت‌دیگر می‌توان ضرورت تجربه بدنمند را برای توسعه معنا پذیرفت و آزمایش فکری دیگری تدارک دید تا بتواند معناداری در سیستم‌های صوری را بررسی کند (Carter, 2007, p 177-180).

آزمایش فکری سرل از این فرض آغاز می‌شود که ممکن است «شکل‌ها» را به‌قدری دست‌کاری کرد که به‌نظر برسد یک سیستم «شکل‌ها» را درک و استدلال می‌کند و به‌طور مناسب پاسخ می‌دهد. از سوی دیگر کارهای اخیر در پردازش زبان طبیعی ادعا می‌کنند سیستم‌هایی که می‌سازند نه‌تنها در زمان اجرا، بلکه در فرایند ساخت نیز سیستم تنها به «شکل» دسترسی دارد، اما در مسئله زبان ارتباط گویندگان با دنیای واقعی محوریت دارد. زبان برای برقراری ارتباط با دنیا واقعیِ (اعم از فیزیکی، اجتماعی و ذهنی) سخنران، استفاده می‌شود. ازاین‌رو فردی که ادعا می‌کند پاسخ‌های معنادار تولید شده است، می‌بایست استدلال کند و بتواند ارتباط این‌گونه پاسخ‌ها را با اطلاعات مربوط به آن جهان اثبات کند. به‌عبارت‌دیگر برای این‌که انسان یا ماشین یک زبان را یاد بگیرد باید مسئله‌ای که هارناد (۱۹۹۰) به‌عنوان «مسئله پایه‌گذاری نماد» مطرح کرده، بتواند حل کند (Harnad, 1990). هارناد به این موضوع اشاره کرده که یک فرد غیرچینی‌زبان ممکن نیست معنای کلمات چینی را تنها از طریق تعاریف در لغت‌نامه [بدون ارتباط آن‌ها با واقعیت خارجی‌شان] یاد بگیرد (Bender and Koller, 2020, p 5188).

همچنین بندر و کولر بدین باورند مدلی که براساس «شکل و ساختار» آموزش دیده، نمی‌تواند معنا را یاد بگیرد، حتی اگر سخت‌افزار مدرن و تکنیک‌های مقیاس‌پذیری داشته باشد و سیستم تبدیل به مدل‌های ارتباطی شوند که قادر به دریافت حجم عظیمی از داده‌ها هستند؛ صرف‌نظر از این‌که عبور از آزمون تورینگ به‌معنای هوشمندی باشد یا نه. آن‌ها استدلال می‌کنند یک سیستمی که بر روی شکل‌ها آموزش دیده، در یک آموزش حساس، شکننده خواهد بود؛ زیرا این سیستم توانایی اتصال سخن خود با دنیای واقعی را ندارد (Ibid, p 5188).

شاید گفته شود مدل‌های زبانی در ظاهر همیشه محتوا و متون مفهوم‌دار و معنادار تولید می‌کنند؛ پس این اشکال وارد نیست که آن‌ها درکی از معنا و مفهوم ندارند، اما با دقت و تأمل مشخص می‌شود عاملی که معناداری و مفهوم‌داربودن این نوع متون را درک می‌کند، انسان است. به‌بیان‌دیگر این عامل انسانی است که با نسبت‌دادن قصد ارتباطی، ارتباط زبان را با جهان خارجی برقرار می‌کند (Ibid, p 5189).

به‌عبارت‌دیگر اگر سیستم‌ها مبتنی بر زبان برنامه‌نویسی (مانند جاوا) یا زبان طبیعی (نظیر زبان انگلیسی) آموزش دیده‌اند، تنها به «شکل و ساختار» زبان دسترسی دارند نه به مفهوم و معنا و نه ارتباط آن‌ها با گوینده و جهان خارج. به‌بیان‌دیگر سیستمی که مفهوم (معنا) یک زبان برنامه‌نویسی را یاد گرفته است، می‌داند که چگونه «کد» را در آن برنامه اجرا کند. همچنین سیستمی که معنای زبان انسان را آموخته، می‌تواند کارهایی مانند پاسخ به سؤالاتی درباره جهان انجام بدهد، اما انجام وظایفی که در فرایند آموزش از آن‌ها غفلت شده است، غیرممکن است. شکل برنامه‌های جاوا برای سیستمی که ورودی و خروجی این برنامه‌ها را مشاهده نکرده است، اطلاعاتی درباره چگونگی اجرای آن‌ها ارائه نمی‌دهد. همچنین شکل جملات انگلیسی، برای سیستمی که شانسی برای به‌دست‌آوردن رابطه «معنای قراردادی» انگلیسی نداشته و در غیاب هر نوع سیگنالی از قصد ارتباطی، هیچ اطلاعاتی درباره موجودیت‌های خارج از زبان ـ که گوینده ممکن است به آن‌ها اشاره کند ـ ندارد؛ بنابراین یک سیستمی که تنها بر روی «شکل زبان جاوا یا انگلیسی» آموزش دیده است، هیچ راهی برای یادگیری روابط معنایی مربوطه ندارد (Ibid, p 5190).

هرچند از منظر مت کارتر استدلال اتاق چینی نتواند دیدگاه محاسبه‌گرایی را ابطال کنند، اما باز هم این استدلال اهمیت دارد؛ زیرا از نظر وی و دیگر اندیشمندان مزبور تنها صورت و شکل برای درک معنا کافی نیست و درک معنا مستلزم ارتباط با جهان خارج است. این نوع ارتباط فعلاً در سیستم‌ها و مدل‌های زبانی وجود ندارد؛ پس درک معنا نیز توسط مدل‌ها امکان‌پذیر نخواهد بود.

به تحلیل دیگر، آیا مدل‌های زبانی بزرگ امروزی، همان «سیستم‌های نمادین» مورد نظر سرل

هستند؟ پاسخ به این پرسش، دقیقاً محل تحلیل جدید مقاله است. سرل در سال ۱۹۸۰، هوش مصنوعیِ مورد نقد خود را به‌عنوان «سیستم‌هایی که براساس قواعد صوری، نمادها را دست‌کاری می‌کنند» توصیف کرد، اما مدل‌های زبانی بزرگِ امروزی، از چند جهت با آن تصویر تفاوت اساسی دارند:

۱. **یادگیری از داده، نه پیروی از قواعد از پیش نوشته شده:** برخلاف سیستم‌های نمادین، مدل‌های زبانی بزرگ از طریق یادگیری آماری از حجم عظیم داده‌های متنی، الگوهای زبانی را کشف می‌کنند. این امر پرسش از اتاق چینی را در سطحی تازه مطرح می‌کند: آیا «یادگیری» خودِ قواعد، وضعیت را نسبت به «در دست داشتن» و «فهمِ» قواعد از پیش نوشته شده تغییر می‌دهد؟

۲. **معماری ترنسفورمر و توجه (Attention):** مدل‌های جدید با مکانیزم «توجه»، روابطِ درون‌متنیِ پیچیده‌ای را مدل‌سازی می‌کنند که در سیستم‌های نمادی وجود نداشت. این امر باعث شده است برخی مدعی شوند که این مدل‌ها «درکی» از بافت و سیاق دارند، اما تحلیل مقاله نشان می‌دهد که این «درک» نیز در سطح نحوی-آماری باقی می‌ماند و به «دربارگی» و ارجاع به جهان خارج نمی‌انجامد.

۳. **پاسخ به نقد «فهمِ سیستمی»:** یکی از مشهورترین پاسخ‌ها به استدلالِ سرل این بوده که هرچند فرد داخل اتاق چینی را نمی‌فهمد، اما خودِ اتاق به‌عنوان یک کل، چینی را می‌فهمد. در مقالهٔ حاضر، با استناد به تحلیل بندر و کولر، نشان داده شده که این پاسخ نیز کارایی خود را در مواجهه با مدل‌های زبانی از دست می‌دهد؛ زیرا «سیستمِ» مورد نظر (مدل زبانی) نیز صرفاً براساس «شکل» آموزش دیده و هیچ راهی برای دست‌یابی به «معنا» ندارد.

با توجه به سه تفاوت بنیادینی که میان مدل‌های زبانی بزرگ امروزی و سیستم‌های نمادین مورد نظر سرل ترسیم شد، می‌توان نتیجه‌گیری کرد که این تفاوت‌ها نه‌تنها استدلال اتاق چینی را بی‌اعتبار نمی‌سازند، بلکه آن را به سطحی عمیق‌تر و هشداردهنده‌تر ارتقا می‌دهند. به‌بیان دقیق‌تر، اگرچه ال‌ال‌ام‌ها به‌جای پیروی از قواعد برنامه‌نویسی، از طریق یادگیریِ آماری به کشف الگوهای زبانی می‌پردازند و با مکانیزمِ «توجه» روابطِ درون‌متنیِ پیچیده‌ای را مدل‌سازی می‌کنند، اما این پیشرفت‌ها همگی در سطحِ «شکل» (فرم) باقی می‌مانند و هیچ‌کدام به «مفهوم» (معنا) راه نمی‌یابند؛ زیرا یادگیری آماری، صرفاً به بهینه‌سازیِ همبستگی‌هایِ نحوی می‌انجامد و مکانیزمِ توجه نیز هرچند بافتِ زبانی را غنی‌تر بازنمایی می‌کند، اما هیچ‌گاه به «ارجاع به جهانِ خارج» و

«دربارگی و حیث التفاتی» دست نمی‌یابد. ازاین‌رو پاسخِ مشهور به سرل مبنی بر این‌که «خودِ اتاق به‌عنوان یک کل می‌فهمد» نیز در مواجهه با این سیستم‌هایِ آماری، کاراییِ خود را از دست می‌دهد؛ زیرا «سیستمِ» مورد نظر (همان مدل زبانی) نیز صرفاً براساس «شکل» آموزش دیده و هیچ راهی برای دست‌یابی به «معنا» ندارد. ازاین‌رو افزایشِ داده و پارامتر، هرگز نمی‌تواند پلِ کیفیِ لازم برای گذار از «دست‌کاری هوشمندانه نمادها» به «درک واقعی معنا» را فراهم آورد؛ بنابراین استدلال سرل نه‌تنها در عصرِ ال‌ال‌ام‌ها منسوخ نشده، بلکه به‌عنوان نقدی بنیادین درباره ماهیتِ صرفاً شکلیِ این مدل‌ها، بیش از پیش خود را نمایان می‌سازد.

نتیجه

پژوهش حاضر با هدف واکاوی امکان «درک معنا» در مدل‌های زبانی بزرگ، با تمرکز بر تمایز میان «شکل» و «مفهوم» و با بهره‌گیری از استدلال فلسفی «اتاق چینی» جان سرل، به بررسی یکی از بنیادی‌ترین مسائل در فلسفه هوش مصنوعی و زبان‌شناسی محاسباتی پرداخت. یافته‌های این تحقیق نشان می‌دهد که اگرچه مدل‌های زبانی بزرگ در تولید متون منسجم، پاسخ‌گویی به پرسش‌ها و شبیه‌سازی رفتارهای زبانی انسان به سطحی قابل توجه از کارایی رسیده‌اند، اما این توانایی‌ها لزوماً مستلزم «درک معنا» به‌معنای فلسفی و شناختی آن نیستند.

تحلیل تمایز میان «شکل» (فرم و نحو) و «مفهوم» (معنا و سمنتیک) نشان داد که مدل‌های زبانی عمدتاً براساس الگوهای آماری و ساختاری زبان، آموزش می‌بینند و دسترسی مستقیمی به «قصد ارتباطی»، «معنای قراردادی» و یا ارجاع به جهان خارج ندارند. ازاین‌رو آنچه به‌عنوان «فهم» در عملکرد این مدل‌ها تلقی می‌شود، بیشتر بازتابی از پردازش نحوی پیچیده و تطبیق الگوهای زبانی است، نه دست‌یابی واقعی به‌معنا.

استدلال اتاق چینی نیز به‌عنوان چارچوبی تحلیلی، به‌خوبی نشان می‌دهد که پردازش نمادها براساس قواعد صوری ـ حتی در سطحی بسیار پیشرفته ـ برای تحقق درک معنا کافی نیست. بااین‌حال بررسی‌های انتقادی، به‌ویژه از منظر مت کارتر نشان داد که این استدلال اگرچه در رد برخی صورت‌بندی‌های ساده‌انگارانه از محاسبه‌گرایی مؤثر است، اما به تنهایی برای نفی کامل امکان تحقق معنا در سیستم‌های مصنوعی کفایت نمی‌کند.

نکته کلیدی که از مجموع مباحث به‌دست می‌آید، نقش اساسی «ارتباط با جهان خارج» و «تجربه بدنمند» در شکل‌گیری معناست. معنا نه صرفاً محصول پردازش درون‌زبانی، بلکه حاصل تعامل میان زبان، ادراک و تجربه زیسته در جهان است. ازاین‌رو فقدان چنین پیوندی در مدل‌های

زبانی کنونی، محدودیتی جدی برای انتساب مفاهیم شناختی مانند «فهم» و «دانستن» به این سیستم‌ها ایجاد می‌کند.

همچنین آنچه در مقاله حاضر ارائه شد، صرفاً بازگویی استدلال سرل نبود، بلکه بازخوانی انتقادی آن در پرتو تحولات بی‌سابقهٔ مدل‌های زبانی بزرگ و واکاوی نسبت این استدلال با معماری‌های ترنسفورمری و شبکه‌های عصبی عمیق است؛ امری که در چهار دههٔ پیشین نه موضوعیت داشت و نه امکان‌پذیر بود. سرل هوش مصنوعیِ مورد نقد خود را به‌عنوان «سیستم‌هایی که براساس قواعد صوری، نمادها را دست‌کاری می‌کنند» توصیف کرد. اما مدل‌های زبانی بزرگِ امروزی، از چندین جهت با آن تصویر تفاوت اساسی دارند. ازاین‌رو تسری استدلال اتاق چینی به مدل‌های زبانی بزرگ، خود نوآوری به‌حساب می‌آید، که این نوع پژوهش در ایران اسلامی مسبوق به سابقه نیست. با توجه به سه تفاوت بنیادینی که میان مدل‌های زبانی بزرگ امروزی و سیستم‌های نمادین مورد نظر سرل ترسیم شد، می‌توان نتیجه‌گیری کرد که این تمایزها نه‌تنها استدلال اتاق چینی را بی‌اعتبار نمی‌سازند، بلکه آن را به سطحی عمیق‌تر ارتقا می‌دهند. هرچند مدل‌های زبانی نوع پیشرفته‌تر از هوش مصنوعی‌های پیشین به‌حساب می‌آیند، اما بنا بر این استدلال، امکان درک معنا چالش بنیادینی است که هوش مصنوعی درون اتاق چینی به‌راحتی نمی‌تواند از آن رهایی یابد.

در نهایت می‌توان نتیجه گرفت که نسبت‌دادن مفاهیم معرفتی به مدل‌های زبانی بزرگ، در بهترین حالت، کاربردی استعاری یا فنی دارد و نباید با معنای عمیق فلسفی این مفاهیم خلط شود. این تمایز برای جلوگیری از اغراق در توانمندی‌های هوش مصنوعی و نیز جهت‌دهی صحیح به پژوهش‌های آینده در حوزه زبان، شناخت و هوش مصنوعی ضروری است.

پیشنهاد می‌شود تحقیقات آتی به‌سوی توسعه مدل‌هایی حرکت کنند که علاوه بر داده‌های زبانی، از تعامل حسی، ادراک چندوجهی و پیوند با جهان واقعی بهره‌مند باشند تا امکان بررسی دقیق‌تر مسئله «معنا» در سیستم‌های مصنوعی فراهم شود.

منابع و مآخذ

1. Bender, Emily M. and Koller, Alexander (2020). “Climbing towards NLU: On Meaning, Form, and Understanding in the Age of Data. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics”. **Association for Computational Linguistics**, Online, 5185–5198.
2. Bender, Emily M. et al (2021)**. On the Dangers of Stochastic Parrots: Can Language Models Be Too Big?**, Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency, Canada (pp. 610–623), New York, NY: ACM.
3. Brown, Tom B., et al, (2020), **Language models are few-shot learners**, in Advances in Neural Information Processing Systems 33: Annual Conference on Neural Information Processing Systems 2020, NeurIPS 2020, December 6-12, virtual, H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, Eds.
4. Carter, Matt (2007). **Minds and Computers An Introduction to the Philosophy of Artificial Intelligence**, Edinburgh: Edinburgh University Press Ltd 22 George Square.
5. Chowdhery, Aakanksha, et al (2023). “Palm: Scaling language modeling with pathways,”. https://doi.org/10.48550/arXiv.2204.02311.
6. Devlin, Jacob, et al (2019). **BERT: Pre-training of deep bidirectional transformers for language under standing**. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers), pp 4171–4186, Minneapolis, Minnesota. Associ ation for Computational Linguistics.
7. Douglas Heaven, Will (2020). “OpenAI’s new language generator GPT-3 is shockingly good—and completely mindless”, **MIT Technology Review**.
8. Dwivedi, Yogesh K. et al (2023). “So what if ChatGPT wrote it? Multidisciplinary perspectives on opportunities, challenges and implications of generative conversational AI for research, practice and policy”, **International Journal of Information Management** ,71, 102642.
9. Fierro, Constanza, et al (2024). “Defining Knowledge: Bridging Epistemology and Large Language Models”, **Proceedings of the 2024 Conference on Empirical Methods in Natural Language Processing**, pp 16096–16111.
10. Floridi , Luciano & Chiriatti, Massimo (2020). “GPT-3: Its Nature, Scope, Limits, and Consequences”, **Minds and Machines**, 30:681–694.
11. Hadi, Muhammad Usman, et al (2023). “Large Language Models: A Comprehensive Survey of its Applications, Challenges, Limitations, and Future Prospects”, DOI:10.36227/techrxiv.23589741.v1.
12. https://blog.google/products/search/search-language-understanding-bert/.
13. https://linguistics.washington.edu/people/emily-m-bender.
14. https://www.nytimes.com/2018/11/18/technology/artificial-intelligence-language.html
15. LeCun, Yann (2022). “A path towards autonomous machine intelligence version 0.9. 2, 2022-06-27”, **Open Review**, vol. 62.
16. McCarthy, John W. (1959). “Programs with common sense. In Proc. Tedding

Conf", **on the Mechanization of Thought Processes**, pp 75–91.
17. Ohsugi, Yasuhito, et al (2019). **A simple but effective method to incorporate multi-turn con text with BERT for conversational machine compre hension**. In Proceedings of the First Workshop on NLP for Conversational AI, pp 11–17, Florence, Italy. Association for Computational Linguistics.
18. Rae, Jack W., et al (2021). "Scaling language models: Methods, analysis & insights from training Gopher". arXiv preprint arXiv:2112.11446, 2021.
19. Saussure, Ferdinand de (1959). **Course in General Linguistics. The Philosophical Society**, Translated by Wade Baskin, New York.
20. Searle, John (1980). "Minds, brains, and programs". **Be havioral and Brain Sciences**, 3(3): 417–457.
21. Shanahan, Murray (2022). "Talking about large language models", **CoRR**, vol. abs/2212.03551
22. Stevan Harnad (1990). "The symbol grounding problem". **Physica D**, 42:335–346.
23. Touvron, Hugo, et al (2023)."Llama: Open and efficient foundation, language models", **CoRR**, arXiv:2302.13971v1 [cs.CL] 27 Feb.