

Large Language Models and Logical Reasoning

Subjects: Computer Science, Artificial Intelligence

Contributor: Robert Friedman

In deep learning, large language models are typically trained on data from a corpus as representative of current knowledge. However, natural language is not an ideal form for the reliable communication of concepts. Instead, formal logical statements are preferable since they are subject to verifiability, reliability, and applicability. Another reason for this preference is that natural language is not designed for an efficient and reliable flow of information and knowledge, but is instead designed as an evolutionary adaptation as formed from a prior set of natural constraints. As a formally structured language, logical statements are also more interpretable. They may be informally constructed in the form of a natural language statement, but a formalized logical statement is expected to follow a stricter set of rules, such as with the use of symbols for representing the logic-based operators that connect multiple simple statements and form verifiable propositions.

Keywords: large language models ; deep learning ; symbolic logic ; propositional logic ; logical reasoning

The large language models of deep learning depend on natural language samples from a large corpus of knowledge ^[1]. However, natural language is a communication form that is derived from biological evolution ^{[2][3][4]} and is a descendant of the other related forms of animal communication, a recent adaptation with roles in sociality, social organization, and general behavior in a natural context ^[5]. It is also not considered a rigorous method for communication of knowledge that includes the attributes of reliability and permanence ^{[6][7]}. Likewise, the common definitions of rhetoric include the act of effective writing—but as an art form, not as a scientific practice ^{[8][9]}. This observation suggests that natural language is limited in its capacity for the construction of reproducible knowledge, a problem studied in philosophy, such as in the literature of epistemology and logic ^[6].

This problem is also studied in the context of deep learning models in computer science ^{[10][11][12][13]}. For instance, Traylor and others ^[11] recorded visual data from experiments for whether the logical operations of negation ^[14] and disjunction ^[15] are learned by neural network models. They showed that these models acquire some generalized knowledge of these tasks, but are not yet reliable for generating the elements of logical reasoning. In another study, Evans and others ^[12] published a data set as a benchmark to test for logical entailment, a subtype of propositional logic that relies on symbolic-based reasoning ^{[16][17]}. These formalized methods of logic may be contrasted against the more informal methods of reasoning. However, the formality of propositional logic offers interpretability on the processes of logical reasoning ^{[18][19][20][21][22][23][24]}.

A strict method of application of propositional logic is to show the symbols of logic as Boolean operators on statements of natural language. An example is negation of a statement ^[14]. Given the statement “human perception is innate”, in a classical context, the negation of this statement is formed by NOT as a Boolean operation on this statement. Therefore, the negation of the statement is “human perception is not innate”. For the case of disjunction in propositional logic, at least in a classical context, two simple statements are joined by the Boolean operator OR ^[15]. Therefore, an example is “human perception is either innate or not innate”. However, other natural languages may represent a formal disjunction using a different method, so it is necessary to contrast the pure forms of ideal logic with its use in common practice ^{[15][24][25]}. Logic is a form of reliable expression of concepts for compensation of vagueness and error in interpretation of statements across the natural languages. Therefore, training a machine system on samples of natural language is expected to reproduce these same errors. However, logic is taught, so it is possible to acquire knowledge of logical reasoning.

The alternative to these formalized approaches is to rely on metaphysical and non-mechanistic definitions of general and logical reasoning. A similar problem exists in the common definitions of “intelligence”. These amorously formed terms stem from concepts that precede the best practices in science and knowledge of the mechanics of cognition and perception. In this case, as for other natural processes, a mechanistic approach to the explanation of phenomena favors experimentation, reproducibility, and scientific rigor. A second alternative is to rely on definitions based on a naive perspective, and, therefore, potential exclusion of the mechanistic basis of natural science and any of the rigor from the methods of philosophy. Since the various schools and teachings of philosophy are not expected to influence common

sense as a population phenomenon, it is likely that common sense is often formed by methods without a component of verifiability [26].

Deep learning and its models are a testbed for verification of information-based phenomena [27][28]. For example, studies have shown that these models are capable of generating machine-readable code from queries in a text-based prompt, such as in the form of natural language requests; therefore, this capability shows that logic is computable by the advanced large language models, as in GPT-4 [28][29]. Bubeck and others [29] further showed that logical reasoning in the GPT-4 model is observed in its solutions for mathematical and general reasoning problems. The higher-order capabilities of these models are often referred to as emergent properties that arise from scaling the model with the inclusion of “big data” sets [30].

It may be that large language models are incorporating samples of propositional logic during the training step, where the data source is a natural language corpus. However, these capabilities may be contrasted against training a pretrained model on a large number of samples with formal statements of propositional logic [25]. The presumption is that tokenization of these samples would better capture higher-order logic than “common speech”. Another method is to connect external tools to the deep learning model using a translation layer that communicates natural language into a machine-readable format [31]. These exchanges between the model and external tools are information that is not incorporated into the neural network itself, but these exchanges may be recorded and added to the model during a later training phase.

References

1. Brants, T.; Popat, A.C.; Xu, P.; Och, F.J.; Dean, J. Large Language Models in Machine Translation. In Proceedings of the 2007 Joint Conference on Empirical Methods in Natural Language Processing and Computational Natural Language Learning (EMNLP-CoNLL), Prague, Czech Republic, 28–30 June 2007; pp. 858–867.
2. Hennig, W. Phylogenetic Systematics. *Annu. Rev. Entomol.* 1965, 10, 97–116.
3. Scott-Phillips, T.C.; Kirby, S. Language evolution in the laboratory. *Trends Cogn. Sci.* 2010, 14, 411–417.
4. Pinker, S.; Bloom, P. Natural language and natural selection. *Behav. Brain Sci.* 1990, 13, 707–727.
5. Friedman, R. Tokenization in the Theory of Knowledge. *Encyclopedia* 2023, 3, 380–386.
6. Waddell, W.W. *The Parmenides of Plato*; James Maclehose and Sons: Glasgow, UK, 1894.
7. Owen, G.E.L. Eleatic Questions. *Class. Q.* 1960, 10, 84–102.
8. Merriam-Webster Dictionary. Available online: <https://www.merriam-webster.com/dictionary/rhetoric> (accessed on 6 April 2023).
9. The Britannica Dictionary. Available online: <https://www.britannica.com/dictionary/rhetoric> (accessed on 11 April 2023).
10. Rae, J.W.; Borgeaud, S.; Cai, T.; Millican, K.; Hoffmann, J.; Song, F.; Aslanides, J.; Henderson, S.; Ring, R.; Young, S.; et al. Scaling Language Models: Methods, Analysis & Insights from Training Gopher. *arXiv* 2021, arXiv:2112.11446.
11. Traylor, A.; Feiman, R.; Pavlick, E. Can Neural Networks Learn Implicit Logic from Physical Reasoning? In Proceedings of the Eleventh International Conference on Learning Representations, Kigali, Rwanda, 1–5 May 2023; (in review). Available online: <https://openreview.net/forum?id=HVoJCRLByVk> (accessed on 12 May 2023).
12. Evans, R.; Saxton, D.; Amos, D.; Kohli, P.; Grefenstette, E. Can Neural Networks Understand Logical Entailment? *arXiv* 2018, arXiv:1802.08535.
13. Shi, S.; Chen, H.; Ma, W.; Mao, J.; Zhang, M.; Zhang, Y. Neural Logic Reasoning. In Proceedings of the 29th ACM International Conference on Information & Knowledge Management, Online, 19–23 October 2020; pp. 1365–1374.
14. Horn, L.R.; Wansing, H. Negation. In *The Stanford Encyclopedia of Philosophy*; Stanford University: Stanford, CA, USA, 2015; Available online: <https://plato.stanford.edu/entries/negation> (accessed on 11 May 2023).
15. Aloni, M. Disjunction. In *The Stanford Encyclopedia of Philosophy*; Stanford University: Stanford, CA, USA, 2016; Available online: <https://plato.stanford.edu/entries/disjunction> (accessed on 11 May 2023).
16. Boole, G. *The Mathematical Analysis of Logic, Being an Essay towards a Calculus of Deductive Reasoning*; Macmillan, Barclay, & Macmillan: London, UK, 1847.
17. Leibniz, G.W. *De Progressione Dyadica Pars I*. 1679. In *Herrn von Leibniz’ Rechnung mit Null und Einz*; Hochstetter, E., Greve, H.-J., Eds.; Siemens Aktiengesellschaft: Berlin, Germany, 1966.

18. Klement, K.C. Propositional Logic. Internet Encyclopedia of Philosophy. Available online: <https://iep.utm.edu/propositional-logic-sentential-logic> (accessed on 12 April 2023).
19. Russell, S. Unifying Logic and Probability. *Commun. ACM* 2015, 58, 88–97.
20. Braine, M.D.; Reiser, B.J.; Romain, B. Some Empirical Justification for a Theory of Natural Propositional Logic. *Psychol. Learn. Motiv.* 1984, 18, 313–371.
21. Garcez, A.D.A.; Gori, M.; Lamb, L.C.; Serafini, L.; Spranger, M.; Tran, S.N. Neural-Symbolic Computing: An Effective Methodology for Principled Integration of Machine Learning and Reasoning. *arXiv* 2019, arXiv:1905.06088.
22. Yang, Y.; Zhuang, Y.; Pan, Y. Multiple knowledge representation for big data artificial intelligence: Framework, applications, and case studies. *Front. Inf. Technol. Electron. Eng.* 2021, 22, 1551–1558.
23. Liang, P.; Potts, C. Bringing machine learning and compositional semantics together. *Annu. Rev. Linguist.* 2015, 1, 355–376.
24. Hitzler, P.; Eberhart, A.; Ebrahimi, M.; Sarker, M.K.; Zhou, L. Neuro-symbolic approaches in artificial intelligence. *Natl. Sci. Rev.* 2022, 9, nwac035.
25. De Raedt, L.; Dumancic, S.; Manhaeve, R.; Marra, G. From Statistical Relational to Neuro-Symbolic Artificial Intelligence. *arXiv* 2020, arXiv:2003.08316.
26. Kant, I. Critique of Pure Reason; Weigelt, M., Translator; Penguin Classics: London, UK, 2003.
27. Friedman, R. A Perspective on Information Optimality in a Neural Circuit and Other Biological Systems. *Signals* 2022, 3, 410–427.
28. Vaswani, A.; Shazeer, N.; Parmar, N.; Uszkoreit, J.; Jones, L.; Gomez, A.N.; Kaiser, L.; Polosukhin, I. Attention is all you need. *Adv. Neural Inf. Process. Syst.* 2017, 30, 1–11.
29. Bubeck, S.; Chandrasekaran, V.; Eldan, R.; Gehrke, J.; Horvitz, E.; Kamar, E.; Lee, P.; Lee, Y.T.; Li, Y.; Lundberg, S.; et al. Sparks of Artificial General Intelligence: Early experiments with GPT-4. *arXiv* 2023, arXiv:2303.12712.
30. Wei, J.; Tay, Y.; Bommasani, R.; Raffel, C.; Zoph, B.; Borgeaud, S.; Yogatama, D.; Bosma, M.; Zhou, D.; Metzler, D.; et al. Emergent Abilities of Large Language Models. *arXiv* 2022, arXiv:2206.07682.
31. Schick, T.; Dwivedi-Yu, J.; Dessì, R.; Raileanu, R.; Lomeli, M.; Zettlemoyer, L.; Cancedda, N.; Scialom, T. Toolformer: Language models can teach themselves to use tools. *arXiv* 2023, arXiv:2302.04761.

Retrieved from <https://encyclopedia.pub/entry/history/show/101825>