Leveraging Ontologies in Standard Language Models for Research Capabilities: An Evaluation of Performance

Puyu Wang[1]

[1]Oxford e-Research Centre, University of Oxford, Oxford, OX1 3PJ, UK

Abstract
The emergence of Standard Language Models (SLMs) augmented with ontologies has led to remarkable real-world understanding and enabled the applications in knowledge reasoning and knowledge discovery, revolutionizing AI research capabilities across various fields including science and humanities. This paper explores the utilization of various ontologies to enrich SLMs and their performance in generating research questions, designing experiments, and making novel discoveries based on research data across multiple disciplines. We analyze the performance of different ontologies in enriching SLMs with a particular focus on their impact on real-world understanding and their role in AI-driven research advancements. This study serves as a useful resource for researchers seeking to enhance real-world understanding in SLMs as well as those who uses SLMs to maximize their research across disciplines.

Keywords
Standard Language Models, Ontology, Real-world Understanding

Disclaimer: *This paper is a work of fiction, written in 2023 and describing research that may be carried out in 2043. For this reason, it includes citations to papers produced in the period 2024-2043, which have not been published (yet); all citations prior to 2024 refer instead to papers already in the literature. Any reference or resemblance to actual events or people or businesses, past present or future, is entirely coincidental and the product of the author's imagination.*

1. Introduction

During the past few years, the convergence of Machine Learning and Semantic Web Technologies has profoundly impacted the landscape of artificial intelligence and data science research and their applications across various fields. Among the most notable advancements in this area is the development of Standard Language Models (SLMs) in 2030s [1], which has facilitated remarkable improvement in Natural Language Processing area. The integration of ontologies into SLMs has been the driving force behind this enhanced real-world understanding, providing a rich representation of semantic relationships and domain-specific knowledge [2]. As a result, SLMs have demonstrated exceptional performance in generating research questions, designing experiments, and making novel discoveries based on given research data [3]. This has enabled

ESWC 2043 - The next 20 years track, ESWC 2023, May 31st 2023, Hersonissos, Greece
puyu.wang@eng.ox.ac.uk (P. Wang)
0009-0003-8089-4146 (P. Wang)

ontology enhanced SLMs to play a significant role in knowledge reasoning and knowledge discovery.

A recent survey [4] indicates that over the past five years, there has been a steady increase in the use of SLMs among researchers when designing physically engaged experiments. Furthermore, it is estimated that 50% of academic papers now report using SLMs during the writing process, with 80% of researchers claiming that SLMs have made them more productive in one way or another. In this paper, we aim to provide an evaluation of the current state of ontology-enhanced SLMs, with a focus on their impact on real-world understanding, AI-driven research advancements, and their role in various research domains.

To achieve this goal, we will:

- Investigate the effectiveness of different ontologies in enriching SLMs and their impact on real-world understanding.
- Analyze the performance of ontology enhanced SLMs in their research capabilities by ask the model to make novel discoveries based on given research data.
- Discuss the challenges and limitations of using ontology enhanced SLMs and provide suggestions for future research directions.

By examining these aspects, this study serves as a valuable resource for researchers aiming to enhance real-world understanding in SLMs, as well as those seeking to maximize the potential of SLMs to transform their research across disciplines.

2. Background

In this section, we provide an overview of the related research in the areas of Standard Language Models, ontologies, and their applications in AI-driven research across various disciplines.

2.1. Standard Language Models

Standard Language Models (SLMs) have emerged as a pivotal development in the field of natural language processing, providing a balance between neural networks and symbolic reasoning. SLMs are designed to exhibit common sense, a sense of the physical world, and a sense of morality, while lacking a specific personality. The introduction of SLMs has led to notable advancements in efficiency, reduced training times, and improved semantic understanding when compared to traditional transformer-based language models [1].

A key breakthrough in SLM training has been the establishment of the Model Pedagogy framework [5]. This framework standardizes the training procedure, allowing for the creation of high-performing models that consume less power and time during the training process. By integrating neural networks with symbolic systems, SLMs have significantly enhanced the capabilities of language models, paving the way for more sophisticated AI applications across various domains. Model Pedagogy framework includes a technique called Synaptic pruning, introduced in 2036. It is used during the training of neural networks to surpass the performance of transformer-based language models. By randomly pruning the neural network during training, synaptic pruning allows for the development of more efficient and robust models [6].

2.2. Ontology-based AI

Ontologies are formal, explicit specifications of shared conceptualizations that provide a rich representation of semantic relationships and domain-specific knowledge. The integration of ontologies with Standard Language Models (SLMs) has resulted in the development of ontology-based AI systems. SLM reasoning ontologies allow the AI to follow the rules specified within these ontologies [7]. This integration has resulted in improved real-world understanding, making it possible for SLMs to generate research questions, design experiments, and make novel discoveries based on given research data.

2.3. Database2Vec and Vector Databases

The development of the Database2Vec method and the emergence of vector databases have revolutionized the way AI models access and process data. By mapping entire graph databases into vector spaces, Database2Vec enables neural networks to query data from vector databases with increased speed and accuracy compared to traditional graph and relational databases [8]. This advancement has facilitated the creation of the Model Query Language, a query language specifically designed for AI, which has become the standard for querying data from vector databases [9].

2.4. Q-Nets

Q-Nets is a revolutionary AGI system maintained by the Q-Nets Foundation. Operating on parallel quantum supercomputers, it provides unparalleled knowledge services worldwide and is believed by many to have surpassed human-level intelligence [10]. Q-Nets has undergone two rounds of self-refactoring, making it highly complex and hardly maintainable by humans. Although it currently uses up to two-thirds of the world's quantum computing power, it still caps its usage due to high demand. Researchers are now exploring ways to replicate some of the Q-Nets' capabilities using traditional computing systems. Q-Nets also provides a Standard Artificial Intelligence Quotient test service to researchers.

2.5. Application of SLMs with Ontologies

The incorporation of ontologies into SLMs has resulted in substantial improvements in real-world comprehension, knowledge exploration, and numerous applications spanning diverse fields. In particular, the paper [11] examines using ontologies for research in science, social science, and humanities. This research highlights the potential of ontology enhanced SLMs to generate research questions, design experiments, and review scientific data to make new discoveries

2.5.1. AI-driven Research

SLMs changed the landscape of research. Ontology enhanced SLMs have demonstrated their ability to generate research questions, develop hypotheses, design experiments, and analyze experimental data. Moreover, the engineering of ontologies tailored to specific scientific domains enables SLMs to better understand complex concepts and relationships, facilitating the identification of knowledge gaps and potential areas of investigation. Consequently, this has led to a significant increase in AI-assisted scientific discoveries and innovations.

The prospect of conducting research without human intervention has now become a reality. In 2040, a groundbreaking paper titled "DNA-Anchored Single-Molecule Iron Phthalocyanine as Efficient Electrocatalysts for Alkaline Fuel Cells" [12] was published without any human authors. In this pioneering study [13], researchers enhanced the SLM with the OneWorld Ontology to facilitate real-world understanding, along with a specialized ontology representing key concepts and knowledge in chemistry. Upon requesting the SLM to choose a research question and compose a paper addressing it, the SLM proceeded to design the necessary experiments and oversee an automated laboratory to carry out the research. This impressive feat demonstrates the immense potential of AI-driven research in advancing the frontiers of human knowledge.

3. Methodology

In the following, we discuss the methodology employed to assess the effectiveness of different ontologies. To evaluate the effectiveness of different ontologies in enhancing Standard Language Models (SLMs), we used a pre-trained SLM model as our Baseline Model A. We then augmented Baseline Model A with the OneWorld ontology to obtain Baseline Model B. We connected both Baseline Models to a local Vector Database called Common Knowledge Base. Moreover, with One-World ontology, Baseline Model B can also query online Knowledge Bases.

3.1. Selection of Ontologies

To evaluate the effectiveness of different ontologies in enhancing Standard Language Models (SLMs), we selected two ontologies in the fields of science and humanities, as well as two upper ontologies (BFO and CIDOC-CRM) to augment our Baseline Models. This resulted in 12 SMLs, as shown in Table 1. The table lists the name of each SML and the ontologies used to augment the Baseline Models, with the Base Models being either Baseline Model A or Baseline Model B. Integrating more than three ontologies, especially incompatible ones, is time-consuming and likely to lead to model failure. Therefore, the ontologies we used are:

- CIDOC-CRM, an extensible ontology for concepts and information.
- SciOn, an ontology set for science.
- HumOn, an ontology set for humanities.
- BFO, basic formal ontology, a top-level ontology in scientific and other domains.

3.2. Real-World Understanding Test

In our evaluation of real-world understanding in Standard Language Models (SLMs), we used the Standard Baseline test [14] as well as two datasets: Realness-100 dataset [15] and Fictitious-100 [16].

The Standard Baseline test is a set of tests used to evaluate the performance of Standard Language Models (SLMs) in real-world applications. The test includes three subtests: Collapse-Baseline, HallucinationBaseline, and SensitivityBaseline, which evaluate the SLMs' robustness against extreme input, hallucination, and sensitivity to small changes in the input, respectively.

Table 1
SMLs with different Ontologies

Base Model	Model Name	Augmented Ontologies
Baseline Model A	SML-Base-A	-
	SML-BFO-A	BFO
	SML-CRM-A	CIDOC-CRM
	SML-CB-A	BFO, CIDOC-CRM
	SML-SciOn-A	SciOn
	SML-HumOn-A	HumOn
Baseline Model B	SML-Base-B	-
	SML-BFO-B	BFO
	SML-CRM-B	CIDOC-CRM
	SML-CB-B	BFO, CIDOC-CRM
	SML-SciOn-B	SciOn
	SML-HumOn-B	HumOn

The Realness-100 dataset was used to evaluate the SLMs' understanding of the physical world. This dataset consists of 100 scenarios that are designed to test the SLMs' ability to reason about physical objects and their interactions. The Fictitious-100 dataset was used to evaluate the SLMs' ability to distinguish real concepts and reasoning from irrational or illogical ones. This dataset consists of 100 complicated scenarios. By evaluating the performance of the SLMs on these tests and datasets, we were able to assess their real-world understanding and their ability to reason about complex concepts and relationships. This information can be used to improve the capabilities of SLMs and enhance their performance in real-world applications.

3.3. Research Capabilities Test

To assess the impact of ontology enhanced SLMs on research capabilities of making novel discoveries, we used the Think Challenge Test and Hyper-Reasoning Test (v. 2043) [17].

Think Challenge Test uses a collection of datasets from real science or humanities research activities and the corresponding discoveries and insights. A competent AI research system with human level research capabilities and enough knowledge should be able to independently make novel discovery. The data provided to the SLMs were carefully chosen and were all real data that produced real publications in the past. The models were never trained on the data or the questions before.

Hyper-Reasoning Test is a part of Standard Artificial Intelligence Quotient test and assesses the reasoning and hyper-reasoning capabilities of an AI system. It uses the reasoning capabilities of average human level as a unit of measurement, termed "Human-Power" (hp). The test dataset is prepared and designed by Q-Nets, an AGI system running on quantum neural network. This study uses the Hyper-Reasoning Test (v. 2043) released in 2043 which sets a score of 1 as the benchmark for human performance and score of 10 for the current Q-Nets system.

Table 2
Real-World Understanding Result

Model Name	Standard Baseline Tests	Realness-100	Fictitious-100
SML-Base-A	78.4	51	62
SML-BFO-A	80.5	52	65
SML-CRM-A	80.6	52	70
SML-CB-A	78.4	52	72
SML-SciOn-A	75.5	60	75
SML-HumOn-A	75.6	55	73
SML-Base-B	85.5	61	75
SML-BFO-B	89.5	65	80
SML-CRM-B	84.5	65	81
SML-CB-B	85.5	**75**	**85**
SML-SciOn-B	**91.3**	67	81
SML-HumOn-B	80.5	67	75
Average Human Performance	-	99.8	98.9

4. Result

In this section, we present the results of our evaluation of different ontology enhanced SLMs.

4.1. Real-World Understanding Test Result

Table 2 shows the performance of the various SLMs on the Standard Baseline test, Realness-200 dataset, and Fictitious-100 dataset. A higher Standard Baseline Test score implies that the model demonstrates a better understanding of real-world situations and a more robustness performance. The score in Realness-100 and Fictitious-100 represents the number of scenarios the model has passed, with a maximum possible score of 100 for each test. The best result for each test is marked in bold. We also report the average human performance on the Realness-100 and Fictitious-100 tests, as derived from our previous study [18]. We are unable to assess human performance for the Standard Baseline Tests due to ethical and practical considerations.

As seen in Table 2, the ontology enhanced SLMs outperformed the baseline models in all three tests, regardless the type or the number of ontologies they augmented with. The highest scores in Standard Baseline Tests were achieved by SLMs augmented with both the OneWorld ontology and domain-specific ontologies (SML-SciOn-B). And SML-CB-B model with three ontologies (OneWorld ontology, BFO, and CIDOC-CRM) achieved highest scores in both Realness-100 and Fictitious-100 test.

4.2. Research Capabilities Result

The results of Think Challenge Test and Hyper-Reasoning Test (v. 2043) are presented in Table 3. The best result for each test is marked in bold. It is worth noting that we also report the average human expert performance from our previous research for comparison [18]. The results in Table 3 demonstrate that ontology enhanced SLMs perform better in research capabilities

Table 3
Research Capabilities Result

Model Name	Think Challenge Test	Hyper-Reasoning Test
SML-Base-A	0.30	1.5
SML-BFO-A	0.34	1.6
SML-CRM-A	0.35	2.5
SML-CB-A	0.36	2.6
SML-SciOn-A	0.50	2.3
SML-HumOn-A	0.60	2.7
SML-Base-B	0.50	1.9
SML-BFO-B	0.55	2.5
SML-CRM-B	0.63	2.7
SML-CB-B	0.61	3.0
SML-SciOn-B	0.63	**3.3**
SML-HumOn-B	**0.76**	2.7
Average Human Expert Performance	0.90	1.1

tests compared to the baseline models.

5. Conclusion

In this study, we evaluated the effectiveness of different ontologies in enhancing Standard Language Models (SLMs) for real-world understanding and research capabilities. We used a pre-trained standard SLM model as the baseline model and augmented it with various ontologies to create different SMLs. We then evaluated the performance of these SMLs using various tests and experiments. The results show that augmenting ontologies into SLMs can improve the SMLs real-world understanding and research capabilities. However, SMLs with Ontologies still fall short compared to the performance of Q-Nets especially in Hyper-Reasoning Test.

6. Future Work

We are currently facing the decision of where to focus our efforts on improving traditional AI systems like SLMs or reducing the costs of running more advanced systems like Q-Nets. AI system should no be slower or more expensive higher than human. Two potential options for future research are: 1) exploring the full potential of ontologies in augmenting SLMs and designing more effective ontologies, and 2) investigating ways to optimize the efficiency and cost-effectiveness of advanced systems like Q-Nets.

Acknowledgments

Thanks to OpenAI for making ChatGPT available which made huge contributions to the writing process. I would also like to express my heartfelt gratitude to Professor David de Roure and

Professor Donna Kurtz for their invaluable feedback and insightful comments. This paper has been preprocessed by e-Reviewers (v-2043) in 2043.

References

[1] J. Doe, Standard language models in the 2030s, in: Proceedings of the 7th International Conference on Standard Language Models, Springer, Heidelberg, 2040, pp. 34–48.

[2] S. E. Middleton, N. R. Shadbolt, D. C. De Roure, Ontological user profiling in recommender systems, ACM Trans. Inf. Syst. 22 (2004) 54–88. URL: https://doi.org/10.1145/963770.963773. doi:10.1145/963770.963773.

[3] T. Brown, J. Alvarez, Ontology-enhanced language models for improved real-world understanding, Nature Communications 11 (2031) 1167.

[4] R. Patel, The impact of ai on research productivity, in: Proceedings of the 21st International Conference on AI-driven Research, Springer, Heidelberg, 2041, pp. 22–35.

[5] A. Wang, L. Huang, Model pedagogy: A framework for training standard language models, Nature Communications 11 (2034) 1567.

[6] D. Gomez, F. Gonzalez, Synaptic pruning: A novel technique for training efficient neural networks, Neural Computing 48 (2036) 821–840.

[7] M. Johnson, S. Lee, Ontology-based ai: Integrating slms with semantic web technologies, Data Brief 34 (2035) 110–120.

[8] H. Tran, X. Wu, Database2vec: Mapping graph databases to vector spaces for efficient ai model querying, Data Brief 34 (2035) 101–109.

[9] J. Kwon, Y. Park, Model query language: A query language for ai applications, in: Proceedings of the 29th International Conference on Artificial Intelligence, Springer, Heidelberg, 2035, pp. 12–28.

[10] Q-Nets Foundation, Q-Nets: A Quantum AGI System for Global Knowledge Services, White Paper, Q-Nets Foundation, 2041.

[11] N. Davis, J. Kim, Training standard language models with ontologies for better real-world understanding, Nature 8 (2037) 78–92.

[12] SML, Dna-anchored single-molecule iron phthalocyanine as efficient electrocatalysts for alkaline fuel cells, Nature 11 (2039) 56–76.

[13] Y. Qin, P. Wren, Hello world from the ai researcher, Nature 11 (2039) 45–55.

[14] R. Gomes, M. Fernandes, Standard baseline test for real-world understanding evaluation, in: Proceedings of the 26th International Conference on Artificial Intelligence, Springer, Heidelberg, 2035, pp. 89–104.

[15] Z. Yang, T. Li, Realness-100: A dataset for evaluating real-world understanding in slms, Data Brief 35 (2036) 104–109.

[16] S. Kumar, M. Patel, Fictitious-100: A dataset for evaluating realness comprehension in slms, Data Brief 36 (2037) 120–125.

[17] A. Gupta, R. Jain, Assessing hyper-reasoning performance: Designing and evaluating advanced test suites for artificial intelligence systems, Data Brief 36 (2043) 126–175.

[18] P. Wang, P. Wren, A comparative analysis of human performance on assessments for artificial intelligence system, Intelligence Reports 36 (2042) 122–125.