

Factors to Consider When Selecting a Large Language Model: A Comparative Analysis

Shreekant Mandvikar

Abstract

As organizations strive to integrate these models into their systems, the pivotal challenge they face is selecting the most appropriate alternative. The task of selecting an appropriate LLM for organizational integration remains complex. This paper presents a comprehensive analysis of factors that should be considered when choosing a LLM, aiming to align the selected model with specific organizational goals. The comparison involves five prominent LLMs: ChatGPT, Bard, Lamma, Hugging Face, and GitHub Copilot. The findings highlight the significance of certain factors in LLM selection. Pre-training data diversity, as observed in ChatGPT and Bard, enhances language coverage and response accuracy. Larger models, like ChatGPT and Bard, exhibit superior comprehension and logical responses due to their extensive parameter count. Training time considerations are crucial, with models such as Bard and Lamma requiring months for training, while Hugging Face and GitHub Copilot offer faster training periods. Language support emerges as a key determinant based on organizational needs. Models like Lamma focus on scientific language, while ChatGPT and Bard emphasize broad language coverage. Enterprise readiness, user data control, and real-time research capabilities are pivotal in decision-making. The study also reveals distinctions in model purposes, API capabilities, user feedback mechanisms, and cloud provider support.

Keywords: Large Language Models, Natural Language Processing, Language Model Selection, Enterprise Readiness

Declarations

Competing interests:

The author declares no competing interests.

© The Author(s). **Open Access** 2023 This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution, and reproduction in any medium or format, as long as appropriate credit is given to the original author(s) and source, a link to the Creative Commons license is provided, and changes are indicated. Unless otherwise stated in a credit line to the source, the photos or other third-party material in this article are covered by the Creative Commons license. If your intended use is not permitted by statutory law or exceeds the permitted usage, you must acquire permission directly from the copyright holder if the material is not included in the article's Creative Commons license.

Introduction

Large language models have transformed the area of natural language processing (NLP) in recent years by improving the capabilities of numerous applications such as chatbots, machine translation, and others. Organizations attempting to implement these models into their systems address the difficult issue of determining the best alternative. To guarantee that the chosen model corresponds with specific objectives and priorities, selecting a Language Model for organizational usage requires a comprehensive review of numerous

parameters. A comprehensive comparison table will be provided, evaluating various aspects such as enterprise readiness, license cost, option to disable user data, real-time research/data capabilities, the purpose of the model, limits on number of calls, API capabilities, user feedback capabilities, source of response (index), support by cloud provider, and the number of customers using each model. By considering these factors, organizations can make an informed decision when selecting a language model. This paper sheds light on the issues businesses should consider when analyzing and comparing five prominent big language models: ChatGPT, Bard, Lamma, Hugging Face, and GitHub copilot.

Table 1. Comparison Table:

Factors	ChatGPT	Bard	Lamma	Hugging Face	GitHub copilot
Pre-trained Data	Wide variety of text sources	Diverse internet text	Numerous scientific papers	Curated datasets and community contributions	Curated and licensed datasets from diverse sources
Model Size	175 billion parameters	9.4 billion parameters	8 billion parameters	Varies (large and small models available)	4 billion parameters
Training Time	Several weeks	Several weeks	Few weeks	Varies	Few weeks
Language Support	Broad range of languages	Extensive supports for multiple languages	Focused on scientific language support	Broad range of languages	Multi-lingual support with regional variations
Enterprise readiness	High	Medium	Medium	High	High
License cost	\$\$	\$	\$\$	\$	\$
option to disable user data	Yes	Yes	Yes	Yes	Yes
Real-time Research/Data Capabilities	High	Low	Medium	High	High
Purpose of the model	Generative AI	Conversational AI	Conversational AI	NLU Development	Code assist
Limits on number of calls	Yes	No	Yes	Yes	No
API capabilities	Extensive	Limited	Extensive	Extensive	Limited
User feedback capabilities	Extensive	Limited	Limited	Extensive	Limited
Source of response	Proprietary	Proprietary	Open Source	Proprietary	Proprietary
Support by cloud provider	Yes	No	No	Yes	N/A
Number of customers	1000+	<100	<100	500+	-

Factors to Consider

Pre-training Data: The diversity and quality of pre-training data substantially influence the language model's capacity to understand and deliver accurate responses. Models such as ChatGPT and Bard make use on a wide range of data sources to provide extensive language coverage.

Model Size: The size of the language model can have a big influence on its performance. Larger models featuring billions of parameters, such as ChatGPT and Bard, frequently display outstanding comprehension and provide more logical responses.

Training Time: Large language models can take a long time to train and require a lot of resources. Organizations should weigh the trade-off between training time and model performance. Models like Bard and Lamma require months of training, but Hugging Face and GitHub copilot can potentially be learned in a matter of weeks.

Language Support: Language support could turn out to be crucial depending on the needs of the company. Lamma prioritizes scientific language, whereas ChatGPT and Bard promote wide language coverage. Hugging Face and GitHub copilot provide versatility by supporting a broad range of languages.

Choosing the best LLM for a firm demands taking into account a number of variables. ChatGpt has strong corporate readiness and the option to deactivate user data, but it has higher license charges. Bard and Lamma provide medium business readiness and the ability to deactivate user data. Hugging Face has strong real-time research capabilities but no cloud provider support. As a free model, GitHub Copilot offers programming help but has restricted API features.

ChatGpt or Hugging Face may be acceptable solutions for organizations looking for a powerful LLM with vast features and good support. Bard and Lamma serve enterprises that value user data privacy but lack enterprise readiness. As a free model, GitHub Copilot may be an appealing option for enterprises with limited resources searching for coding assistance.

Conclusion

Choosing the best broad language model for an organization requires taking into account a variety of specifications. Pre-training data, model size, training duration, and language support all have an impact on a model's performance and applicability for certain use cases. Each model described in this comparison has its own set of advantages and disadvantages. To make an educated selection, organizations should examine and compare these elements depending on their particular requirements. Organizations may employ the potential of NLP to enhance their apps, improve user experiences, and achieve their business goals by adopting an appropriate significant language model. In the end, the selection of LLM is determined by unique corporate objectives, use cases, and financial constraints. Evaluating the parameters included in the comparison table will give enterprises with useful insights, allowing them to choose an LLM that best matches their needs and augments their language processing skills, thus leading to increased operational efficiency.

References

- [1] "What is generative AI?," 19-Jan-2023. [Online]. Available: <https://www.mckinsey.com/featured-insights/mckinsey-explainers/what-is-generative-ai>. [Accessed: 09-June-2023].
- [2] Y. Chang *et al.*, "A Survey on Evaluation of Large Language Models," *arXiv [cs.CL]*, 06-Jul-2023.
- [3] H. Wang, J. Li, H. Wu, E. Hovy, and Y. Sun, "Pre-Trained Language Models and Their Applications," *Proc. Est. Acad. Sci. Eng.*, Sep. 2022.
- [4] D. Bulanda and D. Petrecki, "Generative AI for Developers - Our Comparison," *Grape Up*, 22-May-2023. [Online]. Available: <https://grapeup.com/blog/generative-ai-for-developers-our-comparison/>. [Accessed: 04-July-2023].