

Open Source vs. Paid Large Language Models (LLMs): A Comprehensive Documentation

Introduction

Large Language Models (LLMs) are AI models that can generate human-like text. They are used in a variety of applications, including chatbots, content generation, and more. There are two main types of LLMs: open-source models and paid models. This document provides a comprehensive comparison between these two types of LLMs, highlighting their pros and cons to help your company make an informed decision.

OpenAI LLMs (GPT3.5, GPT 4.0)

Pros:

1. High-Quality Responses: OpenAI LLMs are known for providing high-quality responses, making them ideal for applications where accuracy and quality are paramount.
2. Cost-Effective for Low Usage: For low usage or low-quality inference, OpenAI LLMs can be cost-effective. For instance, 1000 text pages cost between \$0.93 - \$42.
3. Fast Time to Market: OpenAI LLMs require minimal infrastructure and staff specialization, enabling faster product development.

Cons:

1. Data Security: Using OpenAI LLMs may expose sensitive data, posing a potential security risk.
2. High Cost for High Usage: For high usage or high-quality inference, OpenAI LLMs can be prohibitively expensive. For example, 500k text pages can cost between \$7k - \$28k.
3. Vendor Lock-In: Relying on a single vendor for LLMs can lead to vendor lock-in, limiting flexibility.
4. Unclear Licensing for Reselling: The licensing terms for "white label" reselling of OpenAI LLMs are unclear.

FOSS (Free and Open Source Software) LLMs

Pros:

1. Data Security: FOSS LLMs can be hosted on-premises, providing better control over sensitive data.
2. Cost-Effective for High Usage: FOSS LLMs are more cost-effective for high usage, offering a predictable long-term cost.
3. Flexibility: FOSS LLMs offer the flexibility to host the infrastructure on any hardware or service.
4. Improved Performance: Fine-tuned FOSS LLMs should perform better than OpenAI's models.

Cons:

1. Initial Roll-Out Cost: Deploying FOSS LLMs involves a large initial roll-out cost.
2. Lower Quality: The quality of FOSS LLMs is slightly lower than OpenAI's models, but it's improving rapidly.

3. Need for Specialized Staff: Managing LLM training and deployment requires specialized staff.
4. Licensing: It's crucial to use the correct open-source license to avoid potential liability or exposure.

Recommendations

If time to market is a major consideration, it is advisable to initially use OpenAI models. There are open-source libraries available that enable a smooth transition from OpenAI to FOSS LLMs later with minimal cost if the initial framework is correctly architected.

It is also crucial to choose LLMs with the appropriate FOSS license to avoid potential liability or exposure in the future. Currently, the optimal licenses for commercial use are Apache 2.0, followed by the MIT License.

As of this writing, the best FOSS models available under the Apache 2.0 and MIT licenses are:

- Falcon-40B
- LAION Open Assistant
- Nomic.ai Gpt4all
- Databricks Dolly2.0

Conclusion

Choosing between open-source and paid LLMs depends on several factors, including your company's specific needs, budget, and technical capabilities. Both types of LLMs have their advantages and disadvantages, and it's crucial to consider all these factors before making a decision.

Please feel free to reach out if you have any further questions or need additional information.

References

- [Large Language Models (LLMs): Pros and Cons of using Open Source VS OpenAI (paid models)]

(<https://www.linkedin.com/pulse/large-language-models-llms-pros-cons-using-open-vs-paid-khan-cissp>)

- [Log management - Wikipedia](https://en.wikipedia.org/wiki/Log_management_and_intelligence)

- [Choosing an LLM Model | Microsoft Learn](<https://learn.microsoft.com/en-us/semantic-kernel/prompt-engineering/llm-models#:~:text=LLM%20AI%20models%20are%20generally,%2C%20learn%20from%2C%20and%20generate.>)

- [List of Open Sourced Fine-Tuned Large Language Models](<https://medium.com/geekculture/list-of-open-sourced-fine-tuned-large-language-models-llm-8d95a2e0dc76#:~:text=Salesforce%20released%20XGen%3A%20A%207B,so%20it%20is%20open%2Dsource.>)

- [OpenSource Large Language Models for Enterprise Success - Webkul](<https://webkul.com/blog/open-source-large-language-models-for-enterprise/#:~:text=There%20are%20two%20main%20types,customers%20who%20purchase%20a%20license.>)