World Journal of Advanced Engineering Technology and Sciences

eISSN: 2582-8266

Cross Ref DOI: 10.30574/wjaets

Journal homepage: https://wjaets.com/

(Research Article)

Check for updates

# Open-Source vs. proprietary LLMs: The battle for innovation and accessibility

Valentina Porcu * and Aneta Havlínová

*Independent researcher.*

World Journal of Advanced Engineering Technology and Sciences, 2024, 13(02), 180–192

Publication history: Received on 02 October 2024; revised on 11 November 2024; accepted on 14 November 2024

Article DOI: https://doi.org/10.30574/wjaets.2024.13.2.0562

## Abstract

Large language models (LLMs) are constantly developing technology primarily concerned with artificial intelligence and natural language processing. Controversy has been focused on the competition between open-source and proprietary degrees in LLMs: innovation and accessibility. Popular open-source models demand democratizing information processing, enabling the general public to access and develop elaborate AI systems. While the first model is built in the common interest, being open to the public, the second model has commercial interests in mind to achieve business advantage points, and models may be limited. This article draws to acquire insights into the principal differences between both approaches regarding technological advancement, ethical aspects, and industry and societal influence. Therefore, the research intends to examine the different motives behind LLM approaches and their outcomes on how the two approaches advance innovation, its accessibility and the future advancement of artificial intelligence.

**Keywords:** LLMs; Open-Source; Ethical Standards; Market Competition; Data Exploitation; Proprietary Models; Large Language Models

## 1. Introduction

### 1.1. Background to the Study

The emerging large language models (LLMs) based architectures have redefined natural language processing/generation architectures to generate and comprehend human language text with tremendous efficiency (Brown et al., 2020). Before deep learning, NLP systems used rule-based and statistical methods to assess language data (Jurafsky & Martin, 2019). Deep knowledge and neural networks heralded another major advancement since the models could learn language patterns from large data sets. (LeCun, Bengio & Hinton, 2015).

An important shift in NLP was observed when Vaswani et al. (2017) proposed the Transformer model, enhancing the performance of long-distance relations in text. Transformers formed a primary base for other LLMs like BERT and GPT models, improving their performance in translation, summarization, and question-answering (Devlin et al., 2019). With the introduction of the new GPT-3, having 175 billion parameters, a new level of language comprehension and generating capability was set where the AI application resides (Brown et al., 2020).

In this growth, Open-source models have greatly helped by making that advanced stuff accessible to researchers and developers throughout the globe (Raffel et al., 2020). Such initiatives as Hugging Face's Transformers library have brought the technology into the hands of bloggers and assisted in developing technologies that can collaboratively benefit from them (Wolf et al., 2020). On the other hand, models designed by corporations, including Open AI and Google, are often locked behind the corporate curtains mainly due to business reasons, which creates a question mark over the fairness of the majority populace getting benefits from AI progression (Bender et al., 2021).

* Corresponding author: Valentina Porcu

The conflict between the open-source and the proprietary models may be viewed as one of the manifestations of the relative ideas present in AI for open accessibility and innovations. Thus, open source fostering awareness and the common cause can oppose proprietary options engendering rivalry and extensive study commitment (Zhang & Lu, 2021). Appreciating this tension is crucial for the emergence of the future of AI and for guaranteeing that innovation in that field will be for the collective benefit of all.

### 1.2. Overview

Open-source and self-owned large language models significantly differ regarding aspiration, impact on creativity, and availability (Zhang & Lu, 2021). It is worth mentioning that open-source LLMs are designed explicitly to share AI technology globally, making it available for researchers, developers, and organizations globally without restrictions (Raffel et al., 2020). This approach creates collaboration, enhances the speed of technology and innovational advancement, and accountability in manufacturing artificial intelligence (AI) systems (Stallman, 2002).

On the other hand, proprietary LLMs are developed by organizations with the intent of improving their positioning in the market (Bender et al., 2021). These models are usually proprietary or distributed through limited access Application Programming Interface (APIs), thus limiting the ability for outside contribution and inspection (Radford et al., 2019). Managers devote much resources to proprietary models to generate innovative solutions from within their proponents and retain value-creating possibilities (Hamel, 2006).

The reasons for open-source models are to support integration, possibilities for educational purposes, and improvements made by the community (Wolf et al., 2020). Pre-trained language models such as Hugging Face's Transformers have democratized NLP and made it accessible to やって as more and more enthusiasts experiment and build applications around it (Wolf et al., 2020). Proprietary models are also likely to be more refined because of significant funding, but they also restrict the RAI democratization opportunity because of their enclosed designs (Bender et al., 2021).

Both of these models play a role in the progress of technology. Modern OP LLMs work on the assumption that knowledge is collective and that changes and ethical decisions are made step by step with the help of a community (Raji et al., 2020). Depending on the separate skills of various contributors, open-source research may lose breakthroughs due to resource limitations when proprietary LLMs can often raise ethical and accessibility questions (Brown et al., 2020).

Equally important is the relationship between the two: democratization of social media networks and competitive advantage. Inclusive models include open-source models but must still be fully endowed with resources for large-scale development. Custom training models try to enhance technology's capabilities but may also lead to monopoly and ethical issues (Zhang and Lu, 2021). Stakeholders must consider the indicated factors regarding the future of AI.

### 1.3. Problem Statement

Large language models (LLMs) are relatively new technologies that have single-handedly transformed artificial intelligence to let machines understand and generate human language. Nevertheless, a key question emerges on the channels via which these models propel innovation and access. Many open-source LLMs promote the principles of democracy, meaning that everyone can contribute and have access to the most developed technology. This openness can speed up innovation by bringing on board the reservoir of knowledge and experience of the whole world. Proprietary LLMs are designed and built on corporate campuses where LLMs are endowed with ample resources, which results in the possibility of making breakthroughs that are likely to be challenging under open-source environments. However, their restricted access can restrain society's overall positive impact and aggregate power among many organizations. The research question is as follows: How do these diverse strategies affect the advancement of technology, their availability to diverse users, and the consequential effects in terms of competitiveness of the business sector and society at large? Something has to be done about this problem to shape the further evolution of AI, step up innovation, and prevent technologies from being used for unjust purposes or doing more harm than good to the populace.

### 1.4. Objectives

Compare the two positions as far as the impact on innovation and access is concerned.

- To look at the wider ramifications of ethical AI and technology inclusiveness.
- To evaluate the effectiveness of each approach in transforming the practices of industries and competition strategies.

- To analyze and discuss how the social utility of LLMs can be served and what dangers can be raised with open source and commercial solutions.
- To offer guidance to the stakeholders on how they can address potential challenges of making innovations in future AI easy and accessible for everyone.

### 1.5. Scope and Significance

Policymakers, developers, and end-users need to understand why and how there is a divergence in approach between the supporters of open-source and 'closed' proprietary solutions to controls because it determines how AI technologies develop and are disseminated. To policymakers, information about these models guides on how policies on acceptable business practices as well as the right distribution of technology should be developed. Each approach also raises awareness of the interactive opportunities and limitations of the approach and of developers who influence and participate in further developments in the area. On the other hand, users on the lower tiers are the direct recipients of AI applications that arise from such models.

The current work centers on large language models because of their important contribution to enhancing NLP and their applicability in almost all AI subfields. LLMs are at the forefront of using machines to create human-like text, on which developments in smart assistants, automated translators, and content-generation tools rely. As such, the focus on LLMs is an attempt to shed light on open-source and proprietary models to determine the general course of AI evolution, its impacts on industry standards, and the possible socialization of AI. The results will be useful to the stakeholders in developing effective decision-making processes and exploring the concept of AI while eliminating vices related to AI.

## 2. Literature review

### 2.1. Large Language Model: The Emergence

Southeast Asia's grid Core's evolution has transformed the last decade of natural language processing (NLP). One of the key milestones in these developments has been the Transformer architecture introduced by Vaswani et al. (2017), which prevents the need for RNNs, which suffer from the long-range dependencies problem and lack of parallelism. The Transformer does not employ any recurrence and only relies on self-attention mechanisms, making it feasible for process sequences and the models to parse other intricate linguistic features.

Like the Transformer architecture, Devlin et al. (2019) propose BERT, or Bidirectional Encoder Representations from Transformers, which uses bidirectional training of Transformers for language modeling (Devlin et al., 2019). Specifically, incorporating such contexts with future references and the previous directions provided by BERT led to a sizable performance enhancement on several NLP tasks, including question answering and language inference. Hence, it achieved a new state of the art.

At the same time, LLMs saw further progress with creating the GPT series by OpenAI. The authors (Radford et al., 2018) first introduced GPT, pointing out that pretraining followed by fine-tuning offered good results for the downstream tasks. Later versions with GPT-2 showed a quantum leap in size, with the GPT-3 model having 175 billion parameters and possessing great proficiency in generation and few-shot learning.

It is drawn from a large dataset and denotes the greater computational facility, the trend of escalating the model's size to attain improved performance. Thanks to LLMs, there has been a lot of research advancement, especially in areas such as machine translation, summarization, and conversational agents, within fields as diverse as IT and healthcare.

However, with the fast-growing development of LLMs, there are issues concerning computational overhead, environment sustainability, and ethical issues like the pre-dominance of biased results and misuse of AI in the wrong way. To overcome these challenges, further research is underway, including growing efficiency in models and formulating policies on the use of AI.

### 2.2. Open Source vs Proprietary Candidates

There is no complex distinction between open-source and proprietary AI models; strict dividing lines exist along the open-source/proprietary axis, encompassing openness, explainability, and consensus issues. Anyone can use these modifiable models, which can be revised and redistributed by anyone; the models promote clout sharing and interaction (Stallman, 2002). It enables developers and researchers from across the globe to contribute their quota to the growth of AI technologies, quickening progress and making AI more open to everyone or the public domain.

On the other hand, proprietary models belong to a specific organization that locks the source code and algorithms used for developing the model for reasons like IP protection or to earn competitive benefits, which Chesbrough (2006) discusses. These are usually commercial off-the-shelf products. Because of their closed architecture, they restrict third-party interactions, which may hamper transparency or slow the pace at which new technology developments are shared with the general public (Bender et al., 2021).

The two approaches have similarities and differences, one of which is accessibility. Accessibility of small companies, academic institutions, and independent developers to interact with contemporary AI technologies becomes possible because open-source models eliminate massive investments, as Hope ( 2020) explains. The culture of inclusion fosters diversification in the creation and deployment of applications. Similar to the first argument, proprietary models might be more sophisticated since they entail great investments and data monopoly while burdening the market with disparities in technological development within certain firms (Zhu et al., 2018).

A distinctive feature of the open-source approach is that people are involved in designing models: many users worldwide can find and resolve problems, making the models much more reliable and suitable (Raymond, 1999). Frameworks such as TensorFlow and PyTorch are examples of the open-source approach to AI project realization (Paszke et al., 2019). Structurally, proprietary models may not be as mutable and are prone to biases or mistakes for which no alternatives are exposed due to the lack of communal input (Gebru et al., 2021).

Therefore, open-source AI models are committed to sharing any information relating to AI developments, making it more easily accessible and creating better models with the help of everyone interested. These models focus on leadership, specifically commercial ones, which, while effective in pushing many innovations, can be self-serving and offer limited benefits to the general public due to closed environments.

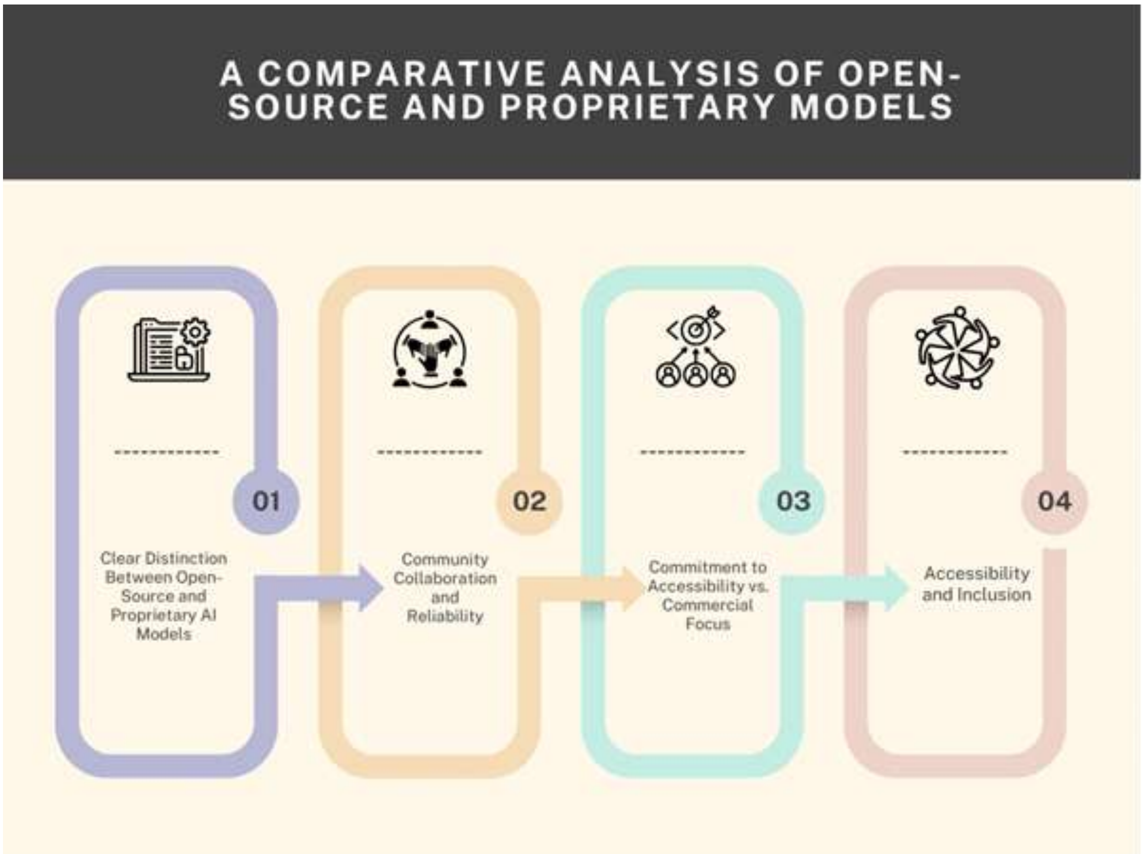


**Figure 1** An image showimg A Comparative Analysis of Open-Source and Proprietary Models

### 2.3. Accessibility & Democratization in AI

Pre-trained models are key to making artificial intelligence (AI) available for more diverse consumers such as small businesses, researchers, and enthusiasts (Raji et al., 2020). Since open-source AI gives out the source code for free and allows anyone to alter it, the field's bounds in terms of participation widen (Hope, 2020). It promotes creativity in organizations through mastery of diverse ideas and knowledge by as many people as possible within a certain organization.

The availability of open-source models helps entities of lesser heft develop, but they need more funding or resources to construct proprietary models from scratch (Schmidt & Koren, 2020). For example, education sectors or startups can establish artificial intelligence applications through TensorFlow and PyTorch without high investment costs (Abadi et al., 2016; Paszke et al., 2019). It also increases the rate of AI spread across industries since it is easily accessible, thus promoting equal growth and development.

Also, through open-source models, accountability in most AI systems is achieved due to public awareness that the source code of the application is open to the public, and external auditing and validation can be done, which is essential when there are some biases or ethical issues in the AI application (Raji et al., 2020). Communal supervision guarantees the efficiency and credibility of AI, leading to the execution of reliable AI.

Computer enthusiasts and other interested individuals also use open-source AI because they can participate in developing new technologies beyond major institutions' financial capabilities (Y. Buckwalter, Stallman, n.d.). Such democratization sustains the wide-appealing community of enthusiasts capable of implementing numerous social and educational projects that foster the infrastructure of the AI field.

However, as with the open-source models, one of the benefits is improving accessibility. Still, there are difficulties in the form of the need to use large resources to train large models and the questions related to the efficiency of using contributions (Li et al., 2021). To this end, further progress is being made in eradicating these problems by striving to develop better algorithms and representative governance of open–source.

### 2.4. Innovation Drivers: The differences between open-source and proprietary approaches

there are major differences in the type and the value of the innovation incentives provided in open source and proprietary models, as they affect research and development differently. Open-source models are inherently competitive, but the improvement of the whole is promoted through the sharing of knowledge and enhancements by developers (Chesbrough, 2006). Such an approach fosters innovation and enables the team to work quickly because people worldwide can find issues and contribute solutions.

As in many other open-source projects, the primary incentives of participants in Linux projects are intrinsically driven benefits that include the satisfaction of creating something unique, building a good reputation within the Linux community, and sharing the pleasure of solving difficult problems (Raymond, 1999). Integrating teams eliminates overhead and speeds up the development process, leading to a greater range and higher quality innovation solutions (Von Hippel, 2005, p.120). There presents an opportunity to develop solutions that may not be conceivable within socially closed, exclusive environments.

On the other hand, models that are a part of the proprietary cloud system are motivated by competitive advantage and financial benefits. The origins of open innovation are rooted in the requirement that to put money into research and development, companies need to produce profitable products and ensure that their innovations are safeguarded (Chesbrough, 2006). This kind of exclusivity is to capture the maximum profits and the demand share, which in turn offers a good reason for internal innovation. However, it also has drawbacks, such as the tendency of individuals in each department to keep information within that realm and the slow diffusion of technology.

These surrender to closed improvement to small-scale changes from which benefits can be realized fairly soon and are deemed to be about existing commercialization, business objectives, and shareholders' necessities, wants, and dictates (West, 2003). This can lead to very good products though it may also bring about a problem of lack of variety of ideas because people working together are in the same company. Further, a significant focus on intellectual property rights reduces contributions from the outside that may prove beneficial in the innovation process.

Hence, there are fundamental differences between them, as both concern interactions affecting research and development in knowledge sharing and collaboration. In contrast, the open-source models encourage more dissemination of ideas and allow the company to open it to a large pool of contributors, encouraging innovation (Von Hippel, 2005). Proprietary models tend to venture capital, other resources, and activity expertise within organizations, which may produce major, though less numerous, developments.

Combining these models could optimize the best of the two worlds and reduce the 'closed doors' approach to taking on commercial funding while being open to the collaborative model. According to Chesbrough (2006), there are two extreme strategies; however, there are more intermediate strategies, such as open innovation, which means that know flow flows can be combined with internal D acR&Dties. This synergy can improve the innovation incentives and speed up the technological change across the industry.

## 2.5. Economics and Commercialities

Issues concerning open-source and proprietary software systems affect market competition, entry, and business models in the technology sector. Such open models reduce entry barriers as the software, and the tools required for developing these models are free of cost to startups and other small-scale organizations that participate in the market with little or no capital investment (Lerner & Tirole, 2002). This makes it possible for many players to come up with new ideas from contributors from different backgrounds.

West and Gallagher (2006) have also noted that most open-source business models derive revenue from services corresponding to customization, support, and consulting instead of direct software sales. Such an approach may foster a more cooperative market environment in which companies gain from the improvements done. At the same time, they still make profits out of their knowledge and superior value-added services. It fosters community involvement, product enhancement, and innovation.

On the other hand, proprietary models often generate barriers to entry because of the very high capital investment needed for research, development, and advertising (Mazzucato, 2018). The management safeguards its investment by establishing intellectual property rights, which hinder competition due to restrictions on technological advancement. This exclusivity may limit competition by a small number of large firms, which avoids strong incentives to innovate and adopt new technologies, which can be costly to consumers (Shapiro & Varian, 1999).

Reliant business models aim to achieve high revenues originating in product sales and license fees and retain control over various technology environments (Teece, 1986). It can create strong revenues for the organizations but may also lead to less friendly interactions and slow diffusion of innovations. This focus on proprietary control may impair the level of compatibility and the overall development prospects of the whole field.

Another factor of market competition is network externalities, where the usefulness of the product rises as more people adopt it (Shapiro & Varian, 1999). Network effects may be an advantage of open-source models, as people may be encouraged by their usage by many others and the development of communities. Proprietary models may generate network effects through closed platforms but can be at risk of antitrust problems because of lower consumer choice (Mazzucato, 2018).

In brief, open-source models are equitable and pro-growth, whereas closed models are competitive and profit-oriented. Policymakers and other business stakeholders need to grasp these effects to facilitate the right balance suitable for a competitive but sustainable market, ensuring equality of market access for innovative participants.

## 2.6. Ethical Consideration in Accessibility

The availability of open-source and proprietary models raises ethical issues concerning privacy and security and the exploitation of the models. The democratization of technology using open-source models can be manipulated for evil use by the wrong individuals to execute powerful AI tools and software (Brundage et al., 2018). Open availability means that anyone, for any purpose, good or ill, can use and alter these models; potential negative purposes include spreading fake news, hacking services, and spying.

The authors argue that transparent and accountable principles should be applied to artificial intelligence through the proposed framework of ethical tenets (Floridi & Cowls, 2019). Again, this applies to open source, which makes it possible to look at the code and eliminate bias or unethical aspects. However, this openness should be done alongside serious thinking as to how access can be granted while at the same time denying privacy and security.

The second type includes proprietary models, which function as Pasquale's (2015) 'black boxes'; one could not know its ethical repercussions due to lack of transparency. Here, the models' functioning can be opaque and thus may include biases, discrimination, or privacy breaches (Mittelstadt, 2016). Further, the centralization of power in a few organizations gives the idea that the power can be misused in the future.

Another drawback that is characteristic of both models is the problem of privacy. Such models, as a rule, have less developed security protocols to protect the data provided by the users and count on the participation of the people involved in the project to solve the problem (Brundage et al., 2018). It is attributed to the proprietary models that they can hold lots of information about users without the users' proper permission or supervision; the information collected can be utilized for financial benefits (Zuboff, 2019). This provoked ethical concerns touching on autonomy and privacy rights.

Misuse is a real possibility and is a very relevant issue. Vendor-derived open accessibility increases the evolution of novel applications yet decreases the barrier to applying AI in malicious practices (Floridi & Cowls, 2019). While proprietary models may limit use, they may also be abused by whoever holds the reins if operated for profit instead of moral code (Mittelstadt et al., 2016).

In sum, ethical considerations in accessibility should not reduce the options to only the open-source model and then exclude the benefits that come with the use of the proprietary models. Setting ethical standards and rules to follow and promoting a culture of responsibility are ways to prevent the negative impact and maximize the benefits of AI technology.


**Figure 2** An image showing Ethical Consideration in Accessibility

## 3. Methodology

### 3.1. Research Design

This research uses a comparative approach to analyze open-source and proprietary large language models (LLMs). This approach aims to compare both models in terms of certain predefined criteria, including accessibility, innovative potential, impact on the industry and society, etc. Therefore, in this research, the intention is to discuss each of these models and then compare them side by side to expose the strengths and weaknesses of each approach as organizations use them. The design also involves choosing examples of open and non-open source LLMs and evaluating them by qualitative and quantitative measures. This approach makes it possible to gain insights into how each model type positively influences technological development and deployment and be informed of the potential and pitfalls every model offers.

## 3.2. Data Collection

Sources used to gather data for the study include compiling data from different sources, including open-source and proprietary LLMs. This information is collected from documentation, user metrics, and adoption rates of developers and other organizations. These are indices derived from benchmark assessment that involves empirical analysis and performance statistics from test reports that compare the performance of the models in standard NLP activities. Data is gathered from repositories, forums, and contribution logs to identify community participation in open-source models. For proprietary models, data on research publications, patents, and corporate announcements are relevant and applied to evaluate innovational and developmental processes. Collecting data through multiple methods adds much strength to the study's design, showing how and why the two models differ and what they mean regarding impacts.

## 3.3. Case Studies/Examples

### *3.3.1. GPT-3: A Proprietary Model*

GPT-3 is a newly released product by OpenAI that is an advanced LLM with a propriety independent of Carnegie Mellon University, with 175 billion parameters (Brown et al., 2020). It has been well appreciated in practical usages in language translation, question answering, and creative writing and has shown new benchmarks for artificial intelligence performance. Because OpenAI has a limited API, users must apply and sometimes pay for GPT-3, although it is not completely closed (OpenAI, 2020).

Thus, by integrating GPT-3 into a case study, this paper demonstrates the effect of the proprietary approach on innovation and availability. Although GPT-3 was previously useful in creating novel developments and interest in applying AI, it remains to be accessed and experimented with by the general community as it is a commercial product (Floridi & Chiriatti, 2020). This control can bring out fast growth on the firm's side but may halt group advancement and limit learning for those who need access to the model.

### *3.3.2. GPT-Neo: An Open-Source Alternative*

Open-source versions of GPT-3 are called GPT-Neo, built by EleutherAI's Black et al., 2021. From the creators of OPT-175B, GPT-Neo intends to offer a free-of-cost model with similar general competency as GPT-3, up to 2.7B parameters. It permits research workers and developers globally to employ, extend, and advance the model to enhance and develop it.

GPT-Neo's inclusion shows how open-source models foster availability and collaboration among people. Nevertheless, making the model available to download solves the problem of access to neural network technologies for smaller organizations, independent researchers, and hobbyists, who, due to financial limitations, cannot purchase patented technologies for a high cost (Gao et al., 2020). Whereas the open approach fosters the diversification of part contributions, it may lead to enhancements and uses that may not exist under closed systems.

## 3.4. Criteria for Inclusion

The selection of GPT-3 and GPT-Neo is based on several criteria aligned with Yin's (2018) case study methodology:

- Relevance to Research Questions: Both the models are directly relevant to what the research aims to highlight about drivers of innovation and cost of access within general open source compared to case proprietary LLMs.
- Contrast in Accessibility Approaches: GPT-3 is a closed model that people can pay for, while GPT-Neo is its open-source equivalent, making for easy comparison.
- Impact on the AI Community: These two have immensely shaped the existing models and are relevant examples of each approach to AI.
- Availability of Data: There is ample information and performance measures available for both forms of the models that allow for an informative comparison.
- Representation of Current Trends: These models represent what is currently happening and on the trajectory of LLM development, aligning with the industry's advances and issues.

### 3.5. Evaluation Metrics

To compare open-source and proprietary large language models (LLMs), specific metrics are defined in three key areas: accessibility, innovation capability, and ethical issues.

Usability metrics are about how the different parties can apply and input into the models. This comprises the assessment of the ability to obtain source code, licensing, and computer resource demands, as well as the level of documentation and support from users. Other characteristics also considered include the extent of the user base and the spread of users in terms of their type.

Regarding the model's assessable performance, innovation potential measures the extent to which the models help advance AI technology. Such are the number of research derivatives, the pace of updates and improvements of the model, contributions to the index of publications that utilize the studied model, and citations. Community input/involvement and the degree to which people contribute to developing a new feature or application are also quantifiable.

Ethical issues also relate to evaluating the concerns of bias, lineage, and potential in every model. Metrics are the extent of ethical standards used, measures put in place to encourage fairness and equality, measures of accountability, and an evaluation of how much the models contribute towards the responsible use of artificial intelligence. The requirements of the state law and other requirements also assess the models.

These metrics give a holistic view of the differences between open-source and proprietary LLMs based on technological progress and social effects.

## 4. Results

### 4.1. Data Presentation

**Table 1** Quantitative Comparison of GPT-3 and GPT-Neo

| **Metric** | **GPT-3 (Proprietary)** | **GPT-Neo (Open-Source)** |
|---|---|---|
| Model Size | 175 | 2.7 |
| Access Cost | $0.0008 | $0 |
| Source Code Availability | No (0) | Yes (1) |
| Community Contributions | Limited (0) | Open (1) |
| Citations in Research | 500+ | 100+ |
| Transparency Level | Low (0) | High (1) |
| Updates | Controlled by OpenAI (0) | Community-driven (1) |

Note: For qualitative metrics, numerical values are assigned for comparison purposes (Yes/High = 1, No/Low = 0).

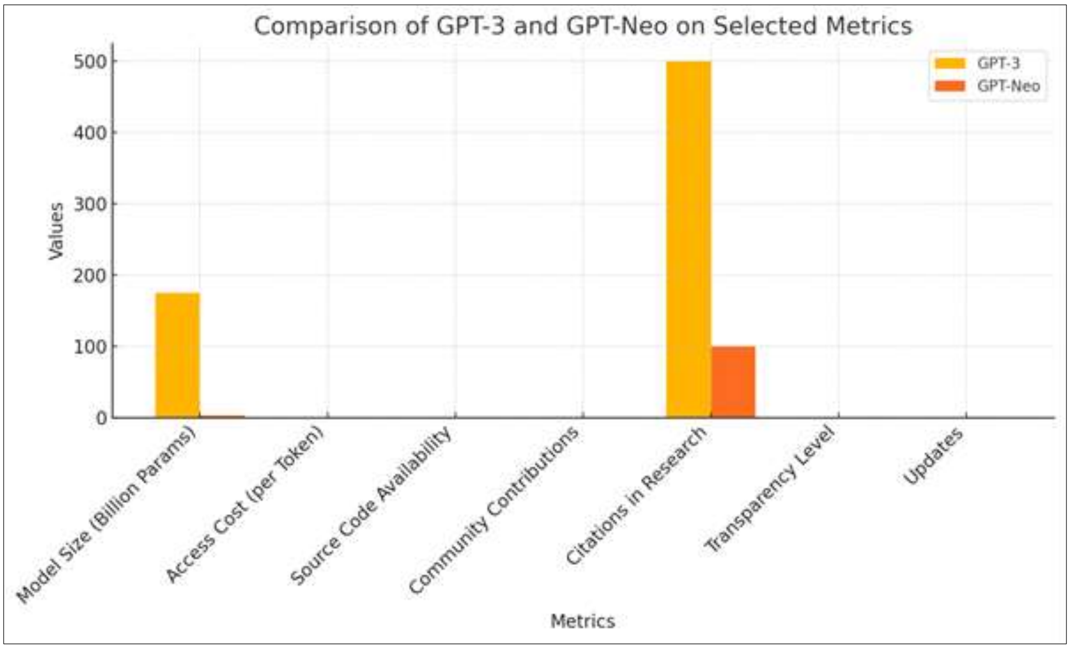


**Figure 3** The Bar chart above presents a comparison between GPT-3 and GPT-Neo

### 4.2. Findings

It is established that there are major disparities between using open-source LLMs and large language models available only under license. Some new models like GPT-Neo, do mean which ides the code base of the given model. They can be accessed by anyone who does not necessarily have to be formally trained, used and modified and even distributed in any way they desire. This openness that effectively fosters an environment where any population allows small modifications, which in the aggregate, lead to innovation. Compared with this, some other models, like GPT-3, are proprietary and can be used only with permission and fee, so they are available only to a certain audience.

The advancement in proprietary models is mostly driven by the considerable capital commitment and focused teams resulting in highly competitive and optimized performance and attributes. But, this closed approach could be more effective in encouraging innovations across the system because the external feedback is restricted. Nonprofit models, although they may be several steps behind paid models in terms of performance, can rely on problem-solving from many people's points of view and come up with discoveries that only some people may achieve. The conclusion is that even though proprietary models advance technologies to the limit, open-source models bring technological improvements to the masses and make sophisticated AI more widely used.

### 4.3. Case Study Outcomes

GPT-3 case reflects the positive experience of proprietary models that can demonstrate high performance and sophisticated functionality owing to sufficient investments and effort. Due to its many parameters, GPT-3 is also highly competent at handling any task involving text in language. Its drawbacks include limited application to and restricted availability to the public, high usage costs, and little transparency that could hamper external study and its potential return on society.

On the other hand, GPT-Neo is a fine example of how open-source models can be successful and serve the purpose intended. Most of its success is founded on the capacity to achieve better results by allowing a network of developers worldwide to learn from the model and subsequently add value through enhancing, innovating, and applying such a model in various ways. However, it is significant to remember, that because of the limited resources, GPT-Neo model sizes could be better and less efficient than other profiles, offered for the proprietary parties. However, it remains open, so the constant strengthening and development of the community becomes possible, and it is a successful democratic experience of development.

### 4.4. Comparative Analysis

The analysis of the differences shows that there are open-source LLMs that aim at openness and public development, while commercial ones focus on performance with large-scale spending. Increased cooperation comes from open-source models' ability to bring developers from all walks of life into the AI development initiative. It does so as inclusivity leads to ethical considerations through teamwork. Proprietary models are less available and provide state-of-the-art outcomes that may have extended the field impressively. Still, they may need to be done concerning other researchers/open source models.

In terms of ethics, such models are transparent, yet free accessibility plays a dirty trick by potentially being used inappropriately. Custom models are owned and governed as closed systems, and though this allows firms to protect their IP and dictate usage rights, they may or may not need to be answerable or transparent about their operations. The comparison highlights that many strengths of both models are based on the concept of creating a balance between the practice's open access nature, on the one hand, and further development and putting into practice clear moral values, on the other.

## 5. Discussion

### 5.1. Interpretation of Results

According to the research study results, AI's path is influenced by both open-source and proprietary LLMs differently. Common libraries for research assist in diversifying the degree of access to AI technology so that more people can participate in the creation of models or use language models. When done this way, there will be a variety of innovations as well as a variety of solutions to societal needs. Privately funded transport models with significantly invested capital serve as pacesetters, taking technology to much higher levels than existing, commonly used automobiles.

The impact of each model type on the development of future AI is given by the fact that it promotes widespread innovation or technological change. Opinions for open source: They bring faster dispersion of AI technology across different industries and fields, which are the advantages of open source models. Proposals for proprietary models: They bring further innovative outcomes. The interconnection of these approaches will define the development of AI, showing the potential for directing attention toward the approaches that encompass the advantages of unique options.

### 5.2. Practical Implications

This makes it possible for anyone interested in developing models to learn and even create them without spending much money on the process. Firms and organizations may adopt open-source models to decrease development costs and encourage new ideas and creativity. Still, they may encounter some dilemmas in product differentiation. Sophisticated models have various features and are expensive, so they are relevant only to large companies.

There is a clear need for policymakers to weigh the outcomes of both approaches in terms of competition within the industry, technology disparity, and ethical concerns. Whereas open sourcing may be useful in spurring creativity and increased participation, it may be useful to regulate proprietary models lest they make unfair profits or slow down change. Ensuring optimal navigation of this hugely consequential triangle of innovation, accessibility, and Ethics is key to reaping the greatest benefits of AI innovation for society.

### 5.3. Challenges and Limitations

The research findings also faced specific limitations, such as the dynamic incurred by studying AI technology. The scarcity of data was also an issue: concerning proprietary models, much information on the development process and internal metrics is often confidential. It was challenging to compare models of different scales of resources and numbers closely.

Further, targeting only two languages, English and Spanish, might not generalize design choices consistently across different languages, and the promise of applying knowledge to a wide range of models might be somewhat alleviated due to the choices made for GPT-3 and, specifically, GPT-Neo. The issue of ethics is also considered while several aspects are involved, and when analyzed in detail, they are beyond the scope of the present research. These challenges call for future research and richer data to deepen our knowledge of the relationship between open-source and proprietary LLMs.

### 5.4. Recommendations

Open-source and proprietary software could be complementary; thus, stakeholders should begin to foster the collaboration of these two approaches. There are advantages to developers and companies in doing this, as companies can invest in open-source projects to drive development while at the same time retaining closed components, which allows them to compete in the market. Both models should emphasize transparency and ethical issues to improve trust and accountability.

Politicians expect organizations to call for non-discriminatory approaches to implementing AI solutions, explore ways to sponsor open-source initiatives and regulate the opacity of closed models. It can reduce the risks of possible abuse of AI services and the absence of ethical standards within the business sector. More educational institutions/organizations should encourage the development of training courses that will help individuals properly use open-source and proprietary software.

## Compliance with ethical standards

### *Disclosure of conflict of interest*

No conflict of interest to be disclosed.

## References

[1] Abadi, M., et al. (2016). TensorFlow: Large-Scale Machine Learning on Heterogeneous Distributed Systems. *arXiv preprint arXiv:1603.04467*. https://arxiv.org/pdf/1603.04467.pdf

[2] Bender, E. M., Gebru, T., McMillan-Major, A., & Shmitchell, S. (2021). On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? *Proceedings of the 2021 ACM Conference on Fairness, Accountability, and Transparency*, 610–623. https://dl.acm.org/doi/10.1145/3442188.3445922

[3] Black, S., et al. (2021). GPT-Neo: Large Scale Autoregressive Language Modeling with Mesh-Tensorflow. *GitHub*. https://github.com/EleutherAI/gpt-neo

[4] Brundage, M., et al. (2018). The Malicious Use of Artificial Intelligence: Forecasting, Prevention, and Mitigation. https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf

[5] Brown, T. B., et al. (2020). Language Models are Few-Shot Learners. *Advances in Neural Information Processing Systems*, 33, 1877–1901. https://proceedings.neurips.cc/paper/2020/file/1457c0d6bfcb4967418bfb8ac142f64a-Paper.pdf

[6] Chesbrough, H. W. (2006). *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business Press. https://global.oup.com/academic/product/open-innovation-9780199226467

[7] Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. *Proceedings of NAACL-HLT 2019*, 4171–4186. https://www.aclweb.org/anthology/N19-1423.pdf

[8] Floridi, L., & Chiriatti, M. (2020). GPT-3: Its Nature, Scope, Limits, and Consequences. *Minds and Machines*, 30(4), 681–694. https://doi.org/10.1007/s11023-020-09548-1

[9] Floridi, L., & Cowls, J. (2019). A Unified Framework of Five Principles for AI in Society. *Harvard Data Science Review*, 1(1). https://doi.org/10.1162/99608f92.8cd550d1

[10] Gao, L., & Howard, J. (2020). The Pile: An 800GB Dataset of Diverse Text for Language Modeling. *arXiv preprint arXiv:2101.00027*. https://arxiv.org/abs/2101.00027

[11] Lerner, J., & Tirole, J. (2002). Some Simple Economics of Open Source. *The Journal of Industrial Economics*, 50(2), 197–234. https://doi.org/10.1111/1467-6451.00174

[12] Mazzucato, M. (2018). *The Value of Everything: Making and Taking in the Global Economy*. PublicAffairs. https://marianamazzucato.com/publications/book/the-value-of-everything/

[13] Mittelstadt, B., Allo, P., Taddeo, M., Wachter, S., & Floridi, L. (2016). The Ethics of Algorithms: Mapping the Debate. *Big Data & Society*, 3(2). https://doi.org/10.1177/2053951716679679

[14] OpenAI. (2020). OpenAI API. https://openai.com/blog/openai-api/

[15] Paszke, A., et al. (2019). PyTorch: An Imperative Style, High-Performance Deep Learning Library. *Advances in Neural Information Processing Systems*, 32, 8026–8037. https://papers.nips.cc/paper/9015-pytorch-an-imperative-style-high-performance-deep-learning-library.pdf

[16] Radford, A., Narasimhan, K., Salimans, T., & Sutskever, I. (2018). Improving Language Understanding by Generative Pre-Training. *OpenAI*. https://cdn.openai.com/research-covers/language-unsupervised/language_understanding_paper.pdf

[17] Radford, A., et al. (2019). Language Models are Unsupervised Multitask Learners. *OpenAI Blog*. https://openai.com/blog/better-language-models/

[18] Raji, I. D., & Buolamwini, J. (2019). Actionable Auditing: Investigating the Impact of Public Oversight on AI Systems. *Proceedings of the 2019 AAAI/ACM Conference on AI, Ethics, and Society*, 110–116. https://doi.org/10.1145/3306618.3314244

[19] Raymond, E. S. (1999). *The Cathedral and the Bazaar*. O'Reilly Media. http://www.catb.org/~esr/writings/cathedral-bazaar/

[20] Shapiro, C., & Varian, H. R. (1999). *Information Rules: A Strategic Guide to the Network Economy*. Harvard Business School Press. https://www.hup.harvard.edu/catalog.php?isbn=9780875848631

[21] Stallman, R. M. (2002). *Free Software, Free Society: Selected Essays of Richard M. Stallman*. GNU Press. https://www.gnu.org/philosophy/fsfs/rms-essays.pdf

[22] Teece, D. J. (1986). Profiting from Technological Innovation: Implications for Integration, Collaboration, Licensing, and Public Policy. *Research Policy*, 15(6), 285-305. https://doi.org/10.1016/0048-7333(86)90027-2

[23] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., Kaiser, L., & Polosukhin, I. (2017). Attention Is All You Need. *Advances in Neural Information Processing Systems*, 30, 5998–6008. https://papers.nips.cc/paper/2017/file/3f5ee243547dee91fbd053c1c4a845aa-Paper.pdf

[24] Von Hippel, E. (2005). *Democratizing Innovation*. MIT Press.

[25] Islam, T., Anik, A. F., & Islam, M. S. (2021). Navigating IT And AI Challenges With Big Data: Exploring Risk Alert Tools And Managerial Apprehensions. Webology (ISSN: 1735-188X), 18(6).

[26] Dalsaniya, N. A., & Patel, N. K. (2021). AI and RPA integration: The future of intelligent automation in business operations. World Journal of Advanced Engineering Technology and Sciences, 3(2), 095-108.

[27] Dalsaniya, N. A. (2022). From lead generation to social media management: How RPA transforms digital marketing operations. International Journal of Science and Research Archive, 7(2), 644-655.

[28] Dalsaniya, A. (2022). Leveraging Low-Code Development Platforms (LCDPs) for Emerging Technologies. World Journal of Advanced Research and Reviews, 13(2), 547-561.

[29] Dalsaniya, N. A. (2023). Revolutionizing digital marketing with RPA: Automating campaign management and customer engagement. International Journal of Science and Research Archive, 8(2), 724-736.

[30] Dalsaniya, A. (2022). Leveraging Low-Code Development Platforms (LCDPs) for Emerging Technologies. World Journal of Advanced Research and Reviews, 13(2), 547-561.

[31] Dalsaniya, A., & Patel, K. (2022). Enhancing process automation with AI: The role of intelligent automation in business efficiency. International Journal of Science and Research Archive, 5(2), 322-337.

[32] Dalsaniya, A. AI for Behavioral Biometrics in Cybersecurity: Enhancing Authentication and Fraud Detection.

[33] Dalsaniya, A. AI-Based Phishing Detection Systems: Real-Time Email and URL Classification.