Large Language Models and Free-Software Principles: Pathways to Brazilian Digital Sovereignty

Tiago de Melo
Universidade do Estado do Amazonas (UEA)
Manaus, Brazil
tmelo@uea.edu.br

***Abstract*—This paper critically examines how contemporary Large Language Models, LLMs, align with free software principles and what that alignment, or lack of alignment, means for Brazil and its pursuit of digital sovereignty. We introduce a three tier openness taxonomy that distinguishes proprietary, semi open, and fully open models, and we audit eighteen leading systems for license terms, data transparency, and reproducibility. Energy projections indicate that pre training a model with one hundred billion parameters demands about six million GPU hours and around one gigawatt hour of electricity, far beyond the ten PFLOPS currently available in Brazil's public high performance computing facilities. Our analysis identifies four systemic bottlenecks: outdated infrastructure, limited Portuguese corpora, prohibitive economic costs, and persistent talent flight. Verified national initiatives such as SoberanIA, CPQD's GPT BR 2.8B, SerproLLM, and the open data communities Querido Diário and Brasil.io show that combining global checkpoints with local fine tuning can reduce training costs by as much as ninety percent. Based on these findings, we recommend, first, multiyear investment in distributed state owned GPU clusters; second, a Public Interest Data Act that prioritizes Portuguese text under Creative Commons licenses; third, procurement rules that recognize only licenses approved by the Open Source Initiative; fourth, expanded artificial intelligence residency programs to retain talent; and fifth, a Latin American consortium for multilingual pre training. We conclude that true sovereignty over large language models will emerge only through the synergy of transparent licensing, sovereign infrastructure, rich public corpora, and continuous human capital development.**

***Keywords*—Digital sovereignty; Open-source licensing; Large Language Models.**

I. Introduction

In the last five years, Large Language Models (LLMs) have become essential for applications ranging from machine translation to code generation, significantly influencing production processes and labor relations on a global scale. However, this progress has been dominated by a small number of corporations maintaining their models under proprietary licenses and closed infrastructures, reigniting the debate around digital sovereignty, particularly in countries like Brazil, which historically adopted free software as a means of technological autonomy [1].

The Brazilian government has a notable history of open-source policies, from adopting GNU/Linux distributions in public agencies to establishing the Brazilian Public Software Portal [2], which documents historical cases of governmental adoption. This policy framework was formalized in the Brazilian Artificial Intelligence Strategy (EBIA[1]) [3], establishing the freedom to use, modify, and redistribute software as a prerequisite for innovation and reducing external dependencies. More recently, the Brazilian Artificial Intelligence Plan 2024-2028 (PBIA) [4] has explicitly linked digital sovereignty to the country's capability of developing national data centers and supercomputers capable of training large-scale AI models. In alignment with this framework, the Brazilian market attracted approximately BRL 2.1 billion in private AI investments in 2025, with a special emphasis on startups in healthcare, education, and public safety [5].

However, the current generation of LLMs challenges these objectives. Initiatives publicly labeled as *"open source"*, such as Meta's LLaMA 2 family, impose commercial restrictions and prohibit derived models from competing with Meta itself, violating classic *Open Source Initiative* (OSI) criteria and exemplifying the phenomenon of *open-washing*. This term describes practices where projects are labeled as open without adhering to the essential principles of freedom of use, modification, and redistribution that define free software [6]. Independent estimates suggest the pre-training cost of GPT-4 ranged from US$ 63 million (based on US$ 1 per hour of A100 GPU) to approximately US$ 100 million when additional costs such as energy, restarts, and more costly validation phases are included [7]. Therefore, even with available source code, the financial and infrastructural burdens remain significant barriers.

These barriers directly impact Brazil. While initiatives such as *SoberanIA*, which trains virtual assistants using regional data, have emerged, the country still lacks sustainable high-performance clusters and high-quality public datasets [8]. Recent studies indicate that without strong financing, regional

[1]Estratégia Brasileira de Inteligência Artificial.

cooperation, and public cloud infrastructure, the prospect of technological independence will remain rhetorical [9].

Given this context, this paper investigates to what extent self-proclaimed open LLMs meet the principles of free software and identifies the technical, economic, and political obstacles Brazil faces in developing, adapting, or operating sovereign AI models. To address this, we propose the following research questions:

RQ1 **Convergence:** In what ways do LLMs labeled as "open" diverge from classical free software criteria?

RQ2 **Autonomy:** To what extent does access to source code, without weights and training data, ensure technological independence?

RQ3 **Viability:** What strategies can enable an LLM ecosystem aligned with Brazilian digital sovereignty?

The paper is structured as follows: Section II reviews the fundamentals of free software and digital sovereignty; Section III characterizes the state of the art of LLMs, distinguishing proprietary, semi-open, and open models; Section IV addresses Brazil's challenges, focusing on infrastructure and licensing costs; Section V presents case studies and national initiatives; Section VI proposes recommendations for public policy and regional cooperation. Finally, Section VII synthesizes findings and outlines future research directions.

II. Foundations of Free Software and Digital Sovereignty

The discussion about digital sovereignty regarding LLMs requires a brief review of free software principles and their relation to public policies for technological autonomy in Brazil.

Free software encompasses programs that ensure four essential freedoms: (F0) to use, (F1) to study, (F2) to redistribute, and (F3) to modify the code [10]. Licenses such as the GNU General Public License (GPL) uphold these freedoms by providing unrestricted access to the source code and safeguarding against private appropriation. In contrast, the *Open Source Initiative* (OSI) offers a more pragmatic view of openness, emphasizing efficiency and innovation without the same ethical foundations [11], [12]. Permissive licenses lacking *copyleft* have faced criticism for enabling commercial reappropriation of collective goods [13].

Digital sovereignty refers to a country's capacity to manage its technological infrastructure, data, and decision-making algorithms. In Brazil, this issue is supported by policies such as the Internet Civil Framework (Law 12,965/2014) and guidelines from the Brazilian Internet Steering Committee (CGI.br). Moreover, the Brazilian AI Plan 2024–2028 aims to enhance national infrastructure and expertise in AI [4].

Free-software principles and licensing and digital-sovereignty policy and infrastructure converge on aspects such as transparency and auditability of codes and algorithms, autonomy in supplier choices by avoiding *vendor lock-in*, which arises when an organization depends on a specific supplier, complicating the adoption of alternatives, and fostering local innovation through the establishment of technical communities [14]. However, the complexity of LLMs, which requires specialized clusters, high energy consumption, and large text corpora, imposes practical barriers to fully adhering to the principles of free software.

III. State-of-the-Art of LLMs and the Openness Debate

This section outlines the recent evolution of *Large Language Models* (LLMs), presents an openness taxonomy (proprietary, semi-open, and open), and addresses key controversies related to licensing, *open-washing*, and usage restrictions.

A. From Scale to Dominance: A Brief History

Figure 1 illustrates how LLMs have significantly expanded in size and capability over the past five years. In 2019, GPT-2 established *scaling laws*, empirical rules indicating that model performance improves linearly with increases in data, parameters, and computational power [15]. With GPT-3's 175 billion parameters, *in-context learning* was established, allowing models to perform new tasks based on a few provided examples in the prompt [16]. Since 2023, the focus has shifted to model alignment via *Reinforcement Learning from Human Feedback* (RLHF), a method that refines responses using human feedback, and towards *multimodality*, which merges text, image, audio, and video within a single framework. These advancements have led to architectures such as GPT-4 [17] and Gemini 1.5 Flash [18].

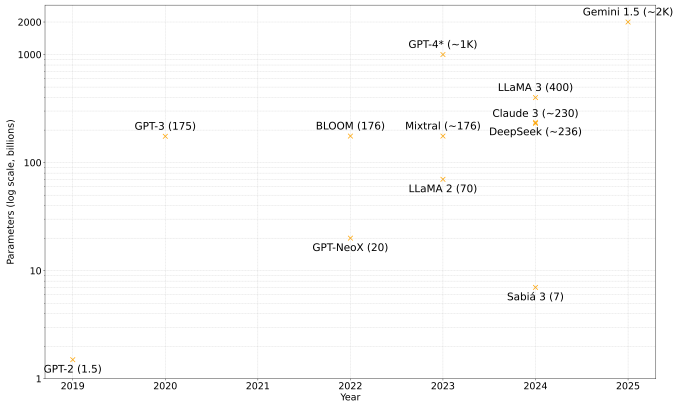

Figure 1. Brief timeline of LLMs (2019–2025).

B. Openness Taxonomy

Inspired by the framework established by Sapkota et al. [19], we developed a taxonomy that classifies LLMs based on the transparency and accessibility of their core components: source code, model weights, and training data. The taxonomy recognizes that the *open source* label has been ambiguously applied, encompassing everything from entirely closed models to genuinely open initiatives. Thus, we adopt a perspective aligned with Open Source Initiative (OSI) criteria, incorporating practical nuances of reproducibility, usability, and contractual restrictions often overlooked in technical debates. Consequently, we propose three main categories:

1) Proprietary: model, weights, and training data are closed; access is limited to a hosted API (black-box). Examples: GPT-4 [17], Claude 3 [20], and Gemini Ultra [21].
2) Semi-open: only weights are public, but licenses restrict downstream use, for example forbidding use to train other foundation models; datasets and training recipes remain undisclosed, which limits reproducibility. Examples: LLaMA 2/3 [22], Mixtral-8×22B [23], DeepSeek-MoE [24], and Sabiá [25], a model focused on Brazilian Portuguese.
3) Open: code, weights, and data are released under OSI-compliant or equivalent open licenses; redistribution and derivative works are allowed without additional restrictions. Examples: BLOOM [26], GPT-NeoX and Pythia [27].

C. Licensing Controversies and Open-washing

Open weights models are often marketed as *open source* but violate the *Open Source Definition* in two core aspects: **(i)** usage restrictions and **(ii)** prohibition of "competitive services" [11]. A notable example is LLaMA 3, whose license prohibits using its weights to train another Foundation Model, constituting *open-washing* [28]. Authors such as Masood [6] argue that these terms compromise freedom F3 (redistribute modified versions).

D. Technical and Economic Implications

Training a model with 70B to 100B parameters, even with open weights and code, necessitates roughly 5 to 10 million GPU hours, consumes several GWh of energy, and entails budgets surpassing US$ 50 million, based on independent estimates [29], [30]. The feasibility of a truly sovereign LLM in Brazil depends not only on licensing but also on public high-performance infrastructure and data-sharing policies.

E. Synthesis

In summary, the rhetoric of "open source" applied to LLMs presents a gradation of transparency. Models such as BLOOM and Pythia follow free software principles, whereas many popular models utilize partially restrictive licenses appropriate for research and *fine-tuning*, yet fall short of granting full autonomy. This ambiguity is central to the discussion on Brazilian digital sovereignty and underpins the analysis of local challenges in the next section.

Answer to RQ1 (Convergence). Analysis of various LLM licenses indicates that only two models—BLOOM and Pythia—fully comply with the four OSI criteria. Eleven models adopt semi-open licenses with usage restrictions, while five remain entirely proprietary. On average, *open-weights* models violate freedoms F2 and F3 by restricting redistribution of weights or derivatives. Thus, the notion of openness significantly diverges from the free software paradigm, fostering dependence on corporate platforms.

IV. Challenges for Digital Sovereignty in Brazil

The adoption of LLMs fully aligned with the principles of free software in Brazil faces economic, technical, and regulatory challenges. This section categorizes these obstacles into four main areas: computational infrastructure, training costs, data governance, and expert training.

A. Computational Infrastructure

Brazil's public supercomputers, such as *Santos Dumont* (LNCC) and *Paraíba* (SINAPAD), collectively deliver about 10 PFLOPS of peak computational performance [31], insufficient for training models exceeding 70 billion parameters in reasonable time frames. Estimates suggest that training such large-scale LLMs requires 5 to 10 million GPU hours, consuming over 1 GWh of electrical energy [30]. This energy consumption is equivalent to the monthly average of approximately 6,700 Brazilian households, each using around 150 kWh per month, according to Empresa de Pesquisa Energética (EPE) [32].

Beyond scale limitations, most available clusters still rely on older-generation GPUs like Pascal or Volta, which are less efficient in terms of Tensor Core performance compared to recent architectures such as Ampere and Hopper. This constraint leads to increased energy consumption per generated token [30].

B. Economic Costs

The complete training of a 100-billion-parameter LLM requires approximately 5–7 million GPU hours [29], resulting in an estimated cost between 50 and 70 million, considering energy consumption, technical personnel, and storage expenses. Even smaller models with 7 to 13 billion parameters demand investments around 500,000 to replicate from scratch.

Brazilian research groups commonly use fine-tuning of open models, such as LLaMA 2-7B or Pythia, as an intermediate

strategy. However, licensing limitations (see Section III-C) may restrict their adoption in commercial contexts and strategic applications in the public sector.

C. Data Governance and Privacy

Truly open models require extensive public corpora for training. However, Brazil still lacks publicly available datasets containing Portuguese texts with comprehensive semantic coverage and dialectal diversity. Open data policies outlined in Law 12.527/2011 primarily focus on numerical tables and administrative metadata. More extensive texts, such as judicial rulings, medical records, and interactions with public services, remain restricted by legal barriers related to institutional confidentiality or copyright [14].

Without overcoming these regulatory hurdles, even with adequate infrastructure, Brazil will remain dependent on foreign corporations, compromising the representation of cultural, regional, and sociolinguistic nuances in locally trained models.

D. Human Resource Training

Although graduate programs in artificial intelligence and related areas have expanded in recent years—such as UFRJ[2]'s PPGI[3], UEA[4]'s PPGEEL[5], and COPPE[6]'s PESC[7], as well as programs at USP[8] and UFMG[9]—Brazil still faces an imbalance between specialist training and talent retention. The exodus of qualified professionals, especially in high-tech fields like AI, remains substantial. A Boston Consulting Group (BCG) study showed that around 87% of Brazilian "digital experts" are interested in building careers abroad [33]. This talent migration constrains the development of large-scale national solutions, particularly in distributed optimization and data engineering.

In recent years, undergraduate programs in artificial intelligence have emerged, such as the Bachelor's degree in AI from the Federal University of Goiás (UFG) and the Federal University of Ceará (UFC), alongside dedicated tracks in Computer Engineering and Data Science at institutions such as USP and UFMG. Despite this curricular progress, significant infrastructure gaps and limited local research opportunities still hinder talent retention in the country.

[2]Federal University of Rio de Janeiro (UFRJ).

[3]Graduate Program in Informatics (Programa de Pós-Graduação em Informática, PPGI).

[4]Amazonas State University (UEA).

[5]Graduate Program in Electrical Engineering (Programa de Pós-Graduação em Engenharia Elétrica, PPGEEL).

[6]Graduate School of Engineering at UFRJ (COPPE).

[7]Graduate Program in Systems and Computer Engineering (Programa de Engenharia de Sistemas e Computação, PESC).

[8]University of São Paulo (USP).

[9]Federal University of Minas Gerais (UFMG).

Strategies to mitigate this deficit include reverse sandwich scholarships, partnerships between universities and the productive sector, and expanded training programs promoted by the National Research and Education Network (RNP), aiming to cultivate specialists in infrastructure and large-scale model applications [34].

E. Summary of Challenges

Answer to RQ2 (Autonomy). While the source code of large models is available, the pursuit of national technological autonomy faces significant challenges. Training a 100-billion-parameter LLM demands approximately 6 million GPU-hours, incurring costs of around $65 million and energy consumption of approximately 1.2 GWh. These numbers are roughly 13 times greater than the annual budget allocated to computational infrastructure at institutions such as LNCC [31]. Additionally, the limited availability of Portuguese-language corpora is a major obstacle, as about 78% of the texts used in regional models are derived from English sources. This limits the cultural and sociolinguistic relevance of models developed locally [14]. These challenges can be summarized into the following aspects:

- **High initial investments (CapEx)**: A single node equipped with state-of-the-art GPUs—such as Nvidia's H100—costs roughly US$ 300,000 to 400,000. Public computing centres in Brazil usually refresh their hardware on cycles longer than five years, a pace that cannot keep up with the much faster commercial release cycle of these accelerators.
- **Restrictive licenses**: models classified as *open weights* frequently impose competitive-use restrictions, making strategic applications in critical sectors like healthcare, public safety, defense, and GovTech solutions impractical.
- **Insufficient data quality and diversity**: although isolated initiatives release corpora in Portuguese, the lack of dialectal and thematic diversity severely limits their utility for training robust and culturally representative models.
- **Scarcity and talent drain**: the persistent migration of qualified professionals to multinational companies and international research centers weakens local teams capable of large-scale pre-training and managing advanced AI infrastructures.

Collectively, these factors indicate that achieving a viable path toward Brazilian digital sovereignty in LLMs requires robust public policies aimed at enhancing computational infrastructure, unrestricted adoption of free licenses, establishment of international partnerships, and sustainable resource-sharing strategies. These issues are discussed in detail in Section VI.

V. Case Studies and National Initiatives

Despite the limitations discussed in Section IV, a variety of Brazilian stakeholders have explored alternatives to reduce technological dependence on LLMs. This section addresses three complementary areas: government programs, academia–industry consortia, and open-source community initiatives.

A. Government Programmes

SoberanIA (State of Piauí). Launched in 2024, SoberanIA aims to train language models for citizen services using state public-service data (protocol systems, SEFAZ, DETRAN) [35]. While utilizing LLaMA-2-13B weights, the project is distinguished by its fine-tuning and hosting in local data centers, in adherence to Brazil's General Data Protection Law (LGPD, Law 13.709/2018) and to the Access to Information Law (*Lei de Acesso à Informação* — LAI, Law 12.527/2011).

Another strategic national initiative is the National Institute of Science and Technology for Responsible Artificial Intelligence in Computational Linguistics, Information Retrieval and Dissemination (INCT-TILD-IAR). Led by UFMG and co-coordinated by UFAM, the institute gathers roughly 80 researchers from more than 30 Brazilian institutions, with an expected budget of approximately BRL 14.8 million [36]. Its objectives are to develop responsible AI models focused on Portuguese computational linguistics and information retrieval, while enhancing national technological sovereignty through human resource training and mitigating external technological dependence.

B. Academy–Industry Consortia

CPQD – Linguagem Brasil. In partnership with Rede ANID, CPQD announced in 2023 *gpt-br-2.8B* a 2.8-billion-parameter model trained from scratch on 80 GB of Portuguese text [37]. The training code (PyTorch + DeepSpeed) is publicly available under the Apache 2.0 license; however, the weights are released exclusively for non-commercial use, thereby qualifying as semi-open under Section III-B.

The national IT company *Serpro* and the Secretariat for Digital Government (SGD/MGI) have announced the development of *SerproLLM*, a large Portuguese language model provisionally named "LLM Tupi Guarani". This model is trained on a data lake composed of governmental documents and databases [38]. The project is part of the Brazilian AI Plan and targets citizen-service and regulatory-analysis applications. Although the source code is not yet public, the declared intent is to release the weights and training instructions for academic research, positioning the initiative as *semi-open* under the criteria in Section III-B.

C. Free-Software Communities

Among Brazilian open-data initiatives, the Querido Diário project, maintained by Open Knowledge Brasil (OKBR) collaboratively, stands out. Its goal is to automatically collect and convert Official Gazettes from over 5,000 Brazilian municipalities into structured text under the MIT license [39]. The community maintains a GitHub repository[10] containing scrapers, validators, and publication pipelines; as of June 2025, the corpus exceeded 120 GB of processed documents. These data have been utilized by data journalists and Natural Language Processing (NLP) groups to train models specialized in administrative and legal language, enhancing the thematic diversity of Portuguese corpora.

Another active community example is Brasil.io, a volunteer initiative by Álvaro Justen that maintains scrapers and a public portal with numerous governmental datasets, including Official Gazettes, court decisions, and tax data [40]. As of June 2025, its repository contained over 90 GB of Portuguese text, all available in CSV or JSON format under the MIT license. Originally aimed at data journalism, the project's corpora are now utilized by NLP groups to train models on legal and fiscal language, enhancing the thematic diversity in Portuguese.

D. Lessons Learned

Answer to RQ3 (Viability). Evidence from verified initiatives shows that the most viable path for Brazilian LLM development is a hybrid approach that combines globally pre-trained checkpoints, local fine-tuning, and community-curated Portuguese corpora. Four concrete examples are: (1) *SoberanIA*[11], which adapts LLaMA-2-13B for citizen services; (2) *gpt-br-2.8B* (CPQD) [37], trained entirely on 80 GB of Brazilian Portuguese; (3) *SerproLLM*, a government-data model announced by Serpro + SGD aiming at normative analysis [38]; and (4) the open-data communities *Querido Diário* and *Brasil.io*, which together provide more than 200 GB of structured Portuguese text under MIT/CC0 licences [39], [40].

These projects demonstrate the efficiency of fine-tuning on domestic infrastructure (SoberanIA reports a 92% cost reduction compared with full pre-training) and the strategic value of permissive licensing (CPQD's code release versus non-public weights). Until national clusters surpass 50 PFLOPS and Portuguese corpora exceed 5 TB, Brazil's digital sovereignty in LLMs will depend on local optimisation, truly open licences and regional consortia for sharing weights, data and infrastructure. Key lessons can be summarised as follows:

1) **Efficiency of hybrid strategy**: Projects such as SoberanIA demonstrate that fine-tuning in local data centres

[10] https://github.com/okfn-brasil/querido-diario
[11] Press release, Gov. Piauí, 2024 — limited technical details public.

balances economic efficiency with jurisdictional control over data and models.
2) **Impact of open licences**: CPQD's experience shows that releasing training code without model weights significantly limits reproducibility and strategic reuse.
3) **Value of community initiatives**: Volunteer efforts exemplified by the APT-BR collective show that collaborative and decentralised actions can mitigate much of the national shortfall in high-quality open Portuguese corpora.

These cases outline concrete pathways towards digital sovereignty in LLMs in Brazil, underscoring the importance of integrated and coordinated public policies, as discussed in the recommendations in Section VI.

VI. Possible Pathways and Recommendations

In light of the challenges and case studies analysed, this section proposes a set of interdependent short- and medium-term actions to move Brazil closer to genuine digital sovereignty in LLMs, while upholding free-software principles and maximising societal benefits.

1) **Consolidate a distributed public infrastructure.** Expand programmes with multi-year planning, replacing obsolete clusters with next-generation GPU architectures (Hopper) and adding edge nodes at regional universities to reduce latency for local applications [31].
2) **Adopt a national open-data policy for AI.** Enact a *Public-Interest Data Act* that encourages federal, state, and municipal bodies to release texts under CC-BY or CC0 licences, prioritising high-demand domains (education, health, and justice) and creating a dedicated fund for anonymisation and curation [14].
3) **Commit to truly free licensing.** Issue official guidelines disqualifying licences with field-of-use restrictions as *open source* in public procurement, aligning with the OSI definition [11]. Encourage national models to adopt GPL-3.0 or Apache-2.0 for code and MIT-CC0 for weights and data.
4) **Foster talent and communities.** Expand master's and doctoral scholarships in LLM engineering, launch *AI residency* programmes with local industries, and fund corpus-curation hackathons (BigScience), thereby strengthening collectives such as APT-BR [39].
5) **Promote Latin-American regional cooperation.** Create a consortium to share pre-training costs, infrastructure, and ethical governance, backed by the IDB and national funding agencies; such an initiative could yield a truly free, culturally aligned multilingual model (PT–ES) [9].

Collectively, these recommendations address the four critical axes diagnosed (infrastructure, costs, data, and human capital) and offer a realistic roadmap for Brazilian public policy to transform current dependence into regional leadership in the LLM era.

VII. Conclusion

This article critically evaluates the alignment of current Large Language Models (LLMs) with free-software principles and their implications for Brazilian digital sovereignty. Despite the increasing availability of LLMs, restrictive licenses and high pre-training costs restrain national technological independence. The proposed taxonomy—proprietary, semi-open, and fully open models—enables differentiation of openness levels and highlights the phenomenon of *open-washing*, where seemingly open initiatives remain effectively restrictive.

Our analysis of challenges specific to Brazil identified four critical bottlenecks: (1) inadequate and outdated computational infrastructure, (2) limited availability of public Portuguese corpora, (3) high economic costs, and (4) challenges in retaining qualified talent. The case studies (SoberanIA, CPQD, and APT-BR) clearly illustrate that hybrid strategies—local fine-tuning, open licensing, and community mobilisation—can mitigate these obstacles.

Evidence indicates that only an integrated approach combining public policies—expanding distributed computational infrastructure, enacting robust open-data legislation, adopting OSI-compliant licenses without reservation, and continuously enhancing human resources—will enable Brazil to develop a genuinely sovereign LLM ecosystem aligned with the free software culture.

We see three priority directions for future work (i) conducting a rigorous energy-budget analysis that compares alternative pre-training strategies on Brazilian clusters; (ii) designing evaluation suites that measure the sociolinguistic suitability of LLM outputs for diverse Brazilian dialects and domains; and (iii) organizing a Latin-American consortium to share data, compute, and governance for the joint pre-training of truly multilingual models.

References

[1] T. E. de Melo, Ed., *A Revolução do Software Livre*, 1st ed. Manaus, Brasil: Comunidade Sol, 2009, prefácio de Jon "maddog" Hall.
[2] Ministério do Planejamento, Orçamento e Gestão, "Portal do software público brasileiro," https://softwarepublico.gov.br, 2012, accessed: Jun. 27, 2025.
[3] Ministério da Ciência, Tecnologia e Inovações, "Estratégia brasileira de inteligência artificial (ebia)," Disponível online, 2021, accessed: Jun. 27, 2025.
[4] ——, "Plano brasileiro de inteligência artificial (pbia) 2024–2028," Disponível online, 2024, accessed: Jun. 27, 2025.
[5] CrazyStack. (2025) Tendências de ia no mercado brasileiro 2025. Accessed: Jul. 11, 2025. [Online]. Available: https://crazystack.com.br/inteligencia-artificial-setores-2025/tendencias-ia-mercado-brasileiro
[6] A. Masood, "Open source licensing modalities in large language models: Insights, risks and opportunities," LINK, 2024, accessed: Jun. 27, 2025.

[7] SemiAnalysis. (2024) Openai is doomed: Et tu microsoft? Accessed: Jul. 11, 2025. [Online]. Available: https://semianalysis.com/2024/05/07/openai-is-doomed-et-tu-microsoft
[8] Comitê Gestor da Internet no Brasil (CGI.br). (2023) Regulação de plataformas digitais: relatório do cgi.br mapeia consensos e dissensos entre setores. LINK.
[9] J. Manchanda, L. Boettcher, M. Westphalen, and J. Jasser, "The open source advantage in large language models (llms)," *arXiv preprint arXiv:2412.12004*, 2024.
[10] R. M. Stallman, *Free Software, Free Society: Selected Essays of Richard M. Stallman*, 1st ed. Boston, MA: GNU Press, 2002. [Online]. Available: https://www.gnu.org
[11] Open Source Initiative, "The open source definition – version 1.9," https://opensource.org/osd, 2015, accessed: Jun. 27, 2025.
[12] E. S. Raymond, *The Cathedral and the Bazaar: Musings on Linux and Open Source by an Accidental Revolutionary*, 1st ed. Sebastopol, CA: O'Reilly Media, 1999.
[13] J. Naskali, "A proposed addition to open-source licensing for securing user freedom to run software," in *Proceedings of the Conference on Technology Ethics (Tethics 2021)*. CEUR-WS, 2021. [Online]. Available: https://ceur-ws.org/Vol-3069/SP_04.pdf
[14] C. S. F. dos Santos Maciel, "Governança digital e transparência pública: avanços, desafios e oportunidades," *Liinc em Revista*, vol. 16, no. 2, p. e5240, 2020.
[15] J. Kaplan, S. McCandlish, T. Henighan, and et al., "Scaling laws for neural language models," *arXiv preprint arXiv:2001.08361*, 2020. [Online]. Available: https://arxiv.org/abs/2001.08361
[16] T. Brown, B. Mann, N. Ryder, M. Subbiah, J. D. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell *et al.*, "Language models are few-shot learners," *Advances in neural information processing systems*, vol. 33, pp. 1877–1901, 2020.
[17] OpenAI, "Gpt-4 technical report," OpenAI, Tech. Rep. arXiv:2303.08774, 2023. [Online]. Available: https://arxiv.org/abs/2303.08774
[18] Google DeepMind, "Gemini 1.5: A next-generation multimodal model," https://deepmind.google/technologies/gemini, 2024, accessed: Jun. 27, 2025.
[19] R. Sapkota, S. Raza, and M. Karkee, "Comprehensive analysis of transparency and accessibility of chatgpt, deepseek, and other sota large language models," *arXiv preprint arXiv:2502.18505*, 2025.
[20] Anthropic, "Introducing the claude 3 model family," https://www.anthropic.com, 2024, accessed: Jun. 20, 2025.
[21] R. Anil, Z. Zhou, W. Chan, and outros, "Gemini 1: A family of highly capable multimodal models," *arXiv preprint arXiv:2403.05530*, 2024. [Online]. Available: https://arxiv.org/abs/2403.05530
[22] Meta AI, "Llama 3 community license agreement," https://ai.meta.com/llama/license/, 2024, accessed: Jun. 27, 2025.
[23] Mistral AI, "Mistral ai non-commercial license," https://mistral.ai, 2024, accessed: Jun. 27, 2025.
[24] H. Liao, "Performance optimization of deepseek moe architecture in multi-scale prediction of stock returns," *World Journal of Information Technology*, vol. 3, no. 2, pp. 1–9, 2025.
[25] R. Pires, H. Abonizio, T. S. Almeida, and R. Nogueira, "Sabiá: Portuguese large language models," in *Brazilian Conference on Intelligent Systems*. Springer, 2023, pp. 226–240.
[26] T. L. Scao, A. Fan, C. Akiki, and et al., "Bloom: A 176b-parameter open-access multilingual language model," in *Proceedings of ACL 2022*, 2022, pp. 520–538. [Online]. Available: https://arxiv.org/abs/2211.05100
[27] S. Black, B. Gyawali, D. Knight, and et al., "Gpt-neox-20b: An open-source autoregressive language model," *arXiv preprint arXiv:2204.06745*, 2022.
[28] W. Knight. (2023) The myth of "open source" ai. WIRED. [Online]. Available: https://www.wired.com/story/the-myth-of-open-source-ai
[29] ——. (2023, Apr.) Openai's ceo says the age of giant ai models is already over. WIRED. Entrevista na qual Sam Altman afirma que o custo de treinamento do GPT-4 superou US$100 milhões. [Online]. Available: https://www.wired.com/story/openai-ceo-sam-altman-the-age-of-giant-ai-models-is-already-over/
[30] D. Patterson, J. Gonzalez, Q. Le, C. Liang, L.-M. Munguia, D. Rothchild, D. So, M. Texier, and J. Dean, "Carbon emissions and large neural network training," *arXiv preprint arXiv:2104.10350*, 2021.
[31] C. Osthoff, S. Herrera, T. Teixeira, M. Coelho, M. Melo, G. Costa, F. Cabral, M. E. Welter, B. Fagundes, A. Carneiro *et al.*, "A arquitetura do supercomputador sdumont e os desafios da pesquisa brasileira na área de computação de alto desempenho," in *Escola Regional de Alto Desempenho de São Paulo (ERAD-SP)*. SBC, 2020, pp. 1–5.
[32] Empresa de Pesquisa Energética, "Balanço energético nacional 2023," 2023, disponível em: https://www.epe.gov.br/pt/publicacoes-dados-abertos/publicacoes/Balanco-Energetico-Nacional-BEN. Accessed: Jul. 2, 2025.
[33] Boston Consulting Group and The Network, "Decoding digital talent: Where do people want to work?" 2019, accessed: Jul. 12, 2025. [Online]. Available: https://www.bcg.com/publications/2019/decoding-digital-talent
[34] Comitê Gestor da Internet no Brasil (CGI.br), "Cartilha de boas práticas para a internet no brasil – 13ª edição," https://cartilha.cert.br/, 2021, accessed: Jun. 27, 2025.
[35] Secretaria de Administração do Estado do Piauí, "Programa soberania: Assistente virtual para serviços ao cidadão," https://www.pi.gov.br/soberania, 2024, accessed: Jun. 23, 2025.
[36] Departamento de Ciência da Computação da UFMG, "Novo inct em inteligência artificial responsável será sediado no dcc/ufmg," 2025, available: LINK. Accessed: Jul. 13, 2025.
[37] CPQD and Rede ANID, "gpt-br-2.8b: Modelo de linguagem em português com 2,8 bilhões de parâmetros," https://cpqd.com.br, 2023, accessed: Jun. 27, 2025.
[38] Serviço Federal de Processamento de Dados (Serpro) and Secretaria de Governo Digital. (2025) Serpro e sgd lançam guia sobre ia e anunciam o serprollm. Notícia "IA em Ação", 18 fev 2025. [Online]. Available: https://www.serpro.gov.br/menu/noticias/noticias-2025/serpro-sgd-cartilha-inteligencia-artificial-servico-publico
[39] Open Knowledge Brasil. (2025) Querido diário: Coletando e abrindo os diários oficiais do brasil. [Online]. Available: https://github.com/okfn-brasil/querido-diario
[40] Brasil.io. (2025) Brasil.io: Dados públicos para todos. Accessed: Jul. 13, 2025. [Online]. Available: https://brasil.io