

# **Evaluating Large Language Model Quality in Resource-Constrained Environments: An Educational Stakeholders' Survey on Accuracy, Completeness, and Readability in Brazil**

**Aristoteles Peixoto Barros<sup>1</sup>, Diego Dermeval<sup>1</sup>,  
Luiz Rodrigues<sup>1,2</sup>**

<sup>1</sup> Center of Excellence in Social Technologies (NEES) –  
Federal University of Alagoas, Brazil  
Maceió – AL – Brazil

<sup>2</sup>Technological Federal University of Paraná (UTFPR) – Apucarana, Brazil  
Apucarana – PR – Brazil

`aristoteles.barros@nees.ufal.br, diego.matos@nees.ufal.br`

`luiz.rodrigues@utfpr.edu.br`

Artificial Intelligence in Education (AIED) has shown potential to personalize learning, automate feedback, and support teachers in classrooms [Chen et al. 2022, Crompton et al. 2024]. However, countries in the Global South still face barriers related to infrastructure, devices, and connectivity, deepening inequalities in educational opportunities [Gasevic et al. 2018, Barron Rodríguez and Cobo 2022]. Lightweight Large Language Models (LLMs), designed to run locally on smartphones, emerge as a viable alternative. This study, derived from a dissertation project [Barros 2024], examines their feasibility and quality in Brazilian education.

The analysis focused on three lightweight LLMs: Gemma, Qwen and RedPajama. Answers to 18 educational questions, based on Bloom's taxonomy and translated into Portuguese, were evaluated by 196 participants from diverse regions. Using Likert-scale questionnaires, evaluators judged accuracy, completeness and readability, following approaches from previous educational technology research [Graesser et al. 2004, Vanlehn 2006]. Unlike studies focused on latency or performance [Monteiro Santos et al. 2024], our emphasis was on perceived response quality.

Results indicate that Qwen and Gemma scored higher in accuracy and completeness, while RedPajama, despite readability, struggled with Portuguese, lowering overall quality. These findings show that LLMs can operate offline in constrained devices with pedagogical relevance, but also highlight the need for language adaptation to ensure equitable impact in contexts such as Brazilian public schools [Isotani et al. 2023].

Implications are significant: lightweight LLMs may democratize access to AIED, enabling teachers and students to benefit from adaptive feedback without internet access [Veloso et al. 2023, Rodrigues et al. 2024]. For researchers, they open opportunities to study trade-offs between efficiency and linguistic quality [Yan et al. 2024]. For policy-makers, they stress the urgency of investing in localized models to mitigate the digital divide. Limitations include reliance on self-reported evaluations, restriction to Earth Sciences questions, and exclusion of Portuguese-specific models. Future work should

extend analyses to other domains, integrate expert validation, and explore optimization techniques to strengthen applicability.

Through the systematic evaluation of lightweight LLMs in offline contexts, this study contributes to the advancement of the AIED unplugged, stressing the urgency of designing educational technologies that are not only effective in fostering learning but also inclusive and accessible to students in underserved regions.

## References

- [Barron Rodríguez and Cobo 2022] Barron Rodríguez, M. R. and Cobo, C. (2022). *COVID-19 and education in the global South: Emergency remote learning solutions with long-term implications*.
- [Barros 2024] Barros, A. P. (2024). Evaluating lightweight llms in educational contexts: Accuracy, completeness and readability. Master's thesis, Universidade Federal de Alagoas.
- [Chen et al. 2022] Chen, X., Zou, D., and Xie, H. e. a. (2022). Two decades of artificial intelligence in education. *Educational Technology & Society*, 25(1):28–47.
- [Crompton et al. 2024] Crompton, H., Jones, M. V., and Burke, D. (2024). Affordances and challenges of artificial intelligence in k-12 education: A systematic review. *Journal of Research on Technology in Education*, 56(3):248–268.
- [Gasevic et al. 2018] Gasevic, D., Prinsloo, P., and Chen, B. e. a. (2018). *Learning Analytics for the Global South*. Foundation for Information Technology Education and Development, Quezon City, Philippines.
- [Graesser et al. 2004] Graesser, A. C., VanLehn, K., and Rosé, C. P. e. a. (2004). Intelligent tutoring systems with conversational dialogue. *AI Magazine*, 22(4):39–51.
- [Isotani et al. 2023] Isotani, S., Bittencourt, I. I., and Chalco, G. C. e. a. (2023). Aied unplugged: Leapfrogging the digital divide to reach the underserved. In *International Conference on Artificial Intelligence in Education*, pages 772–779. Springer.
- [Monteiro Santos et al. 2024] Monteiro Santos, M., Barros, A., and Rodrigues, L. e. a. (2024). Near feasibility, distant practicality: Empirical analysis of deploying and using llms on resource-constrained smartphones. In *Proceedings of the 13th International Conference on Information & Communication Technologies and Development*. ACM.
- [Rodrigues et al. 2024] Rodrigues, L., Guerino, G., and Silva, T. E. V. e. a. (2024). Math-aide: A qualitative study of teachers' perceptions of an its unplugged for underserved regions. *International Journal of Artificial Intelligence in Education*.
- [Vanlehn 2006] Vanlehn, K. (2006). The behavior of tutoring systems. *International Journal of Artificial Intelligence in Education*, 16(3):227–265.
- [Velooso et al. 2023] Velooso, T. E., Chalco Chalco, G., and Rodrigues, L. e. a. (2023). Its unplugged: Leapfrogging the digital divide for teaching numeracy skills in underserved populations. In *Workshop on AI-Augmented Human Tutoring in Math Learning, AIED*.
- [Yan et al. 2024] Yan, L., Sha, L., and Zhao, L. e. a. (2024). Practical and ethical challenges of large language models in education: A systematic scoping review. *British Journal of Educational Technology*, 55(1):90–112.