

# Bridging the Language Divide: Generative AI's Promise for Global Education

**Pulkit Verma**  
MIT CSAIL, Cambridge, USA  
pulkitv@mit.edu

The digital age has brought unprecedented access to knowledge, yet a significant portion of the world's population remains excluded from this wealth of information due to language barriers. This disparity is particularly evident in AI education (and in STEM education too) where English dominates as the primary language of instruction and research [Henry et al., 2021, McDermott, 2023]. However, recent advances in generative AI (GenAI) present a promising solution to this longstanding challenge.

In countries like India and across Africa, where multiple languages coexist with English as a medium of instruction, the language barrier often becomes a significant obstacle to learning. The challenge extends even to developed nations - e.g., in the United States, Spanish-speaking students frequently struggle to access advanced AI and STEM material, leading to initiatives like UCLA's project to increase the availability of neuroscience research in Spanish [McCormack, 2024]. While such isolated efforts are valuable, they require substantial manpower and resources to maintain [Peris-Yague et al., 2024]. The emergence of large language models (LLMs) and multimodal AI systems offers an unprecedented opportunity to address this challenge systematically, as they can democratize knowledge across linguistic boundaries. The recent development of models like Aya [Üstün et al., 2024], capable of processing instructions in 101 languages, represents a significant breakthrough in this field, demonstrating how AI can effectively address the needs of traditionally low-resource languages. The potential impact extends beyond mere text translation. Major platforms like Coursera are already implementing AI-powered translations across multiple languages [Stein, 2023], making higher education more accessible to global audiences. In the multimedia sphere, the technology's capabilities are expanding rapidly, with implications for video content, educational animations, and interactive learning materials.

While concerns about accuracy and reliability of GenAI for low resource languages are valid [Ahuja et al., 2023], they shouldn't overshadow the transformative potential of this technology. The traditional approach of creating educational content in multiple languages has been bottlenecked by the scarcity of bilingual subject matter experts and the high costs associated with translation and localization [Turner et al., 2014]. GenAI offers a practical middle ground: using AI for initial content translation and generation, while employing subject matter experts for verification rather than creation from scratch.

Critical to the success of this technological revolution is the active participation of volunteers and cultural experts. While LLMs can handle the heavy lifting of initial translation, the nuanced understanding of cultural contexts, sensitivities, and local educational needs requires human insight [Hanrahan et al., 2015, Kalfoss, 2019]. The volunteer community plays a vital role in identifying and addressing potential pitfalls, particularly in culturally sensitive material that might be misinterpreted or inappropriately translated by AI systems. Looking ahead to the next 5-10 years, the combination of autonomous translation systems and human validation holds transformative potential for global education. With proper safeguards and community involvement, this approach could revolutionize access to higher education and college-level learning materials. Communities that have historically been marginalized due to language barriers could gain access to world-class educational resources.

This technological solution must be implemented thoughtfully, with careful consideration for cultural nuances and educational quality. The goal isn't merely to translate content but to make knowledge truly accessible and meaningful in local contexts. Through the combined efforts of AI technology and human expertise, we stand at the threshold of creating a more equitable global educational landscape, where language is no longer a barrier to learning and growth.

## References

- Kabir Ahuja, Harshita Diddee, Rishav Hada, Millicent Ochieng, Krithika Ramesh, Prachi Jain, Akshay Nambi, Tanuja Ganu, Sameer Segal, Mohamed Ahmed, Kalika Bali, and Sunayana Sitaram. MEGA: Multilingual evaluation of generative AI. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, 2023.
- Donna Hanrahan, Patriona Sexton, Katrina Hui, Jennifer Teitcher, Jeremy Sugarman, Alex John London, Mark Barnes, James Purpura, and Robert Klitzman. Linguistic and cultural challenges in communication and translation in US-sponsored HIV prevention research in emerging economies. *PLoS one*, 10(7):e0133394, 2015.
- Kaylee R Henry, RK Virk, Lindsay DeMarchi, and Huei Sears. A call to diversify the lingua franca of academic STEM communities. *Journal of Science Policy and Governance*, 18(10):38126, 2021.

Mary Kalfoss. Translation and adaption of questionnaires: A nursing challenge. *SAGE Open Nursing*, 5: 2377960818816810, 2019.

Kayla McCormack. Making STEM accessible through translation. *UCLA Magazine*, February 2024.

Amy McDermott. English is the go-to language of science, but students often do better when taught in more tongues. *Proceedings of the National Academy of Sciences*, 120(40):e2315792120, 2023.

Alba Peris-Yague, Jenna Hartstein, Arielle Hogan, Carla Suhr, and Rafael Romero-Calderón. Increasing accessibility to neuroscience through translation: Going beyond the english language. *eNeuro*, 11(1), 2024.

Marni Baker Stein. Coursera expands AI-powered translations to 17 popular languages. *Coursera Blog*, December 2023.

Anne M Turner, Margo Bergman, Megumu Brownstein, Kate Cole, and Katrin Kirchhoff. A comparison of human and machine translation of health promotion materials for public health practice. *Journal of Public Health Management and Practice*, 20:523–529, 2014.

Ahmet Üstün, Viraat Aryabumi, Zheng Yong, Wei-Yin Ko, Daniel D’souza, Gbemileke Onilude, Neel Bhandari, Shivalika Singh, Hui-Lee Ooi, Amr Kayid, Freddie Vargus, Phil Blunsom, Shayne Longpre, Niklas Muennighoff, Marzieh Fadaee, Julia Kreutzer, and Sara Hooker. Aya model: An instruction finetuned open-access multilingual language model. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, 2024.