

Abstract

Artificial intelligence (AI) systems often reflect biases from economically advanced regions, marginalizing contexts in economically developing regions like Latin America due to imbalanced datasets. This paper examines AI representations of diverse Latin American contexts, revealing disparities between data from economically advanced and developing regions. We highlight how the dominance of English over Spanish, Portuguese, and indigenous languages such as Quechua and Nahuatl perpetuates biases, framing Latin American perspectives through a Western lens. To address this, we introduce a culturally aware dataset rooted in Latin American history and socio-political contexts, challenging Eurocentric models. We evaluate six language models on questions testing cultural context awareness, using a novel Cultural Expressiveness metric, statistical tests, and linguistic analyses. Our findings show that some models better capture Latin American perspectives, while others exhibit significant sentiment misalignment ($p < 0.001$). Fine-tuning Mistral-7B with our dataset improves its cultural expressiveness by 42.9%, advancing equitable AI development. We advocate for equitable AI by prioritizing datasets that reflect Latin American history, indigenous knowledge, and diverse languages, while emphasizing community-centered approaches to amplify marginalized voices.