CARMA 2026

8th Int. Conf. on Advanced Research Methods and Analytics
1-3 July 2026, Valencia, Spain

Quantifying Linguistic Anthropomorphism in Multilingual LLM

Juan Luis Blanco-Guzmán[1], **Jeisson A. Higuera Reina**[2]

[1]Department of Financial Economics and Business Administration, University of Seville, Spain, [2]Department of Marketing & Business Administration, University of Seville, Spain.

How to cite: Blanco-Guzmán, J. L.; Higuera Reina, J. A. (2026). Quantifying Linguistic Anthropomorphism in Multilingual LLM. In: 8th International Conference on Advanced Research Methods and Analytics (CARMA 2026). Valencia, 1-3 July 2026. https://doi.org/10.4995/CARMA2026.2026.22000

Abstract

This study quantifies linguistic anthropomorphism in 25 large language models (LLMs) using the LMSYS Chat 1M dataset (HuggingFace), analyzing over one million conversational turns across seven languages (English, Spanish, Portuguese, Russian, Chinese, German, French). We detect key markers, first-person pronouns, empathetic phrases, emojis, and sentiment polarity, via multilingual dictionaries and regex patterns, revealing significant variations in anthropomorphic expression by model and language. Findings show certain LLMs (e.g., GPT-4, Claude) exhibit higher self-referential and empathetic language, with cultural differences influencing response patterns (e.g., more emojis in Romance languages). These insights highlight risks of overtrust and ethical concerns in AI design, offering metrics for responsible multilingual deployment and user engagement evaluation.

Keywords: *Anthropomorphism; Large Language Models; Multilingual NLP; Sentiment Analysis; Chatbot Analytics; Human-AI Interaction.1. Introduction*

Article in Press

This work is licensed under a Creative Commons License CC BY-NC-SA 4.0
Editorial Universitat Politècnica de València