

Multimodality in multilinguality

Abstract

This is a review of literature in the area of multimodality usage for the multi-language understanding. Multimodality in multilinguality explores the convergence of multimodal and multilingual capabilities in artificial intelligence systems to foster deeper understanding and enhanced adaptability. This review begins by examining foundational work in improving language representation for underrepresented languages with a focus on the available papers. These papers lay a foundation for understanding how multilingual AI models can be extended and optimized. They emphasize the need for innovative embedding techniques, diverse linguistic datasets, and adaptive model architectures to fill the gap between multimodality and multilinguality. This review will further integrate these insights with multimodal models to propose frameworks for deeper and more inclusive AI understanding across modalities and languages.

1 Introduction

The rapid development of large language models (LLMs) has revolutionized the fields of natural language processing (NLP) and artificial intelligence (AI). However, the dominance of English-centric models has created imbalances in the representation of underrepresented languages, limiting the inclusivity and accessibility of these technologies. Multilingual AI seeks to bridge this gap, enabling models to understand and generate text across diverse languages. At the same time, multimodal AI—integrating text, vision, and other modalities—offers an opportunity to achieve deeper perception and broader applicability. The intersection of these domains, multimodality in multilinguality, represents an exciting frontier in AI research.

This review begins by addressing foundational challenges in multilingual representation, focusing on recent advances aimed at improving the inclusion of non-Latin script languages in LLMs. We explore how methods like vocabulary expansion, embedding initialization, and fine-tuning have been employed to enhance the performance of LLMs for underrepresented languages, using Ukrainian as a case study. The papers, "From English-Centric to Effective Bilingual: LLMs with Custom Tokenizers for Underrepresented Languages" [1] and "From Bytes to Borsch: Fine-Tuning Gemma and Mistral for the Ukrainian Language Representation," [2] illustrate the potential of targeted interventions in advancing multilingual models.

Building on these insights, this review aims to connect the progress in multilingual LLMs with multimodal approaches, exploring how AI systems can leverage advancements in both domains to process and understand data across languages and modalities. By addressing these interconnected challenges. This work aims to contribute to the development of more inclusive, adaptive, and perceptive AI systems.

2 Review

2.1 Background

2.2 Related Work

2.3 Methodology

2.4 Results

3 Early Work

4 Conclusion

5 References

- 1 From English-Centric to Effective Bilingual: LLMs with Custom Tokenizers for Underrepresented Languages. arXiv:2410.18836v1 [cs.CL] 24 Oct 2024
- 2 From Bytes to Borsch: Fine-Tuning Gemma and Mistral for the Ukrainian Language Representation. arXiv:2404.09138v1 [cs.CL] 14 Apr 2024