

UC Merced

Proceedings of the Annual Meeting of the Cognitive Science Society

Title

Benchmarking LLMs for Mimicking Child-Caregiver Language in Interaction

Permalink

<https://escholarship.org/uc/item/36c9w8qn>

Journal

Proceedings of the Annual Meeting of the Cognitive Science Society, 47(0)

Authors

Liu, Jing

Fourtassi, Abdellah

Publication Date

2025

Copyright Information

This work is made available under the terms of a Creative Commons Attribution License, available at <https://creativecommons.org/licenses/by/4.0/>

Peer reviewed

Benchmarking LLMs for Mimicking Child-Caregiver Language in Interaction

Jing Liu

ENS, Paris, France

Abdellah Fourtassi

Aix-Marseille University, Marseille, France

Abstract

Child-directed speech (CDS) is characterized by its adaptive nature: Caregivers not only talk to children, but engage in dynamic interactions with them. The adaptive/interactive nature of this type of language is understudied in computational modeling research, particularly given the limited availability of naturalistic data. While recent advances in large language models (LLMs) have demonstrated potential for generating viable synthetic dialogue data in various domains, their ability to capture the dynamics of child-caregiver communication remains unexplored. This paper introduces a systematic framework for evaluating LLMs' capacity to generate developmentally appropriate CDS in interaction, examining both static linguistic features and dynamic conversational patterns. We evaluated state-of-the-art LLMs (GPT-4o and Llama 3) against natural interactions from the CHILDES dataset using both single- and multi-turn testing approaches. In single-turn evaluation, models generated responses to individual child utterances, enabling direct comparison with actual caregiver responses. Multi-turn testing assessed sustained interaction capabilities through simulated child-caregiver dialogues. Our results show that while LLMs can successfully approximate surface-level linguistic patterns after few-shot prompting, they struggle with higher-level communicative aspects, with excessive alignment and reduced diversity compared to natural interactions. Our benchmarking framework elucidates both the potential and limitations of LLMs in generating data that preserves the essential properties of child-caregiver language in interactions.