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 dy- namic interactions with them. The adaptive/interactive nature of this type of language is understudied in computational mod- eling research, particularly given the limited availability of nat- uralistic data. While recent advances in large language models (LLMs) have demonstrated potential for generating viable syn- thetic dialogue data in various domains, their ability to capture the dynamics of child-caregiver communication remains un- explored. This paper introduces a systematic framework for evaluating LLMs' capacity to generate developmentally ap- propriate 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, mod- els generated responses to individual child utterances, enabling direct comparison with actual caregiver responses. Multi-turn testing assessed sustained interaction capabilities through sim- ulated 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 re- duced diversity compared to natural interactions. Our bench- marking framework elucidates both the potential and limita- tions of LLMs in generating data that preserves the essential properties of child-caregiver language in interactions.

In D. Barner, N.R. Bramley, A. Ruggeri and C.M. Walker (Eds.), *Proceedings of the 47th Annual Conference of the Cognitive Science Society*