Can pretrained Large Language Models tell us anything about the learning mechanisms used by children? An analysis based on gender bias

Javier Aguado-Orea
Sheffield Hallam University, Sheffield, United Kingdom

Abstract

Syntactic gender marking can be seen as a fully productive parameter from a strictly grammatical point of view (e.g., McDonald & MacWhinney, 1995). For instance, in English, any verb can potentially receive either a male or female pronoun (i.e., he or she) (e.g., Kennison & Trofe, 2003) or more recently a neutral one (Spayer & Schleef, 2019). Usage, however, paints a different picture. IAT-based research (e.g., Greenwald & Banaji, 1995; Nosek et al., 2002) and reading time (Carreiras et al., 1996) reveal a strong gender bias in adult participants. Although the evidence on young infants is clearly more sparse, it yet points in the same general direction. Gender bias is perceived as a reading-facilitation factor in children aged 8 to 10 (Eilers et al., 2019), despite the representation appearing to be less settled than that observed for adults (Arnold et al., 2019). In conclusion, gender bias appears to be acquired gradually, while its exact nature is yet unknown. The primary objective of this study is to use an alternative estimations to model the possible impact of gender-bias in the input. Using the results of two large language models (LLMs) that have already been trained, OpenAI Chat-GPT 3.5 and Llama-2, the state of the relationship between genders and verbs is evaluated. The results, as predicted, demonstrate a very unproductive use of verbs with the subject pronoun "she" and a significantly more productive use of "he," especially at very low LLM temperatures (i.e., 0.1). Using a parallel set of real language speech samples for adults and children, EsLiPro (Aguado-Orea, 2023) has been used to evaluate the productive usage of specific sets of verbs with either "she" or "he." Findings reveal that samples of very young children (e.g., 3 year olds) had a more productive use of "she" over "he" (i.e., used with more verbs) than do LLMs, despite controlling for sample size and vocabulary. This suggests that AI has a considerably stronger gender bias and, although it throws some possible hints, the core mechanisms might not be fully shared with the real learning process.