Les grands modèles de langage peuvent-ils apprendre comme les enfants ?

Abdellah Fourtassi

Aix-Marseille University, Université de Toulon, CNRS, LIS, Marseille, France

abdellah.FOURTASSI@univ-amu.fr

RÉSUMÉ

Les grands modèles de langage comme ChatGPT ont transformé l'intelligence artificielle, mais nécessitent des quantités massives de données — souvent des centaines de milliards de mots — pour atteindre un haut niveau de performance. En comparaison, les enfants humains acquièrent le langage à partir de seulement quelques millions de mots. Cet écart a suscité une nouvelle ligne de recherche visant à développer des modèles capables d'apprendre à partir d'un volume de données plus proche de celui des enfants. En explorant comment des systèmes peuvent apprendre le langage à partir de données limitées et naturelles, ces travaux offrent des perspectives utiles à la fois pour améliorer les modèles d'IA et pour mieux comprendre l'apprentissage humain. Cette recherche réunit des spécialistes de l'IA et des sciences cognitives, avec un double objectif : rendre les systèmes d'apprentissage plus efficaces en données, et mieux comprendre comment le langage se développe chez l'enfant dans des contextes d'apprentissage écologiques.

ABSTRACT

Can Large Language Models Learn Like Children

Large language models like ChatGPT have transformed the field of artificial intelligence, but they require vast amounts of data—often hundreds of billions of words—to reach high performance. In contrast, human children acquire language from just a few million words. This gap has motivated a new line of research focused on building language models that learn from human-scale input, closer to what children actually experience. By investigating how systems can learn language from limited, naturalistic data, this work offers insights that are valuable both for improving AI and for understanding human learning. It sits at the intersection of artificial intelligence and cognitive theories of language acquisition, bringing together researchers who aim to make learning systems more data-efficient and those who seek to model child language development in ecologically valid settings. This talk will explore the challenges and opportunities of this emerging research agenda and its implications.

MOTS-CLÉS : IA, développement du langage, modélisation, apprentissage, cognition.

KEYWORDS: AI, language development, modeling, learning, cognition.

RÉSUMÉ CONFÉRENCE : **Accepté à IA-ÉDU@CORIA-TALN 2025.**