What's the deal with large language models? A Comparative Evolutionary Perspective

Molly Flaherty[*1], Christine Cuskley[2]

[*]Corresponding Author: moflaherty@davidson.edu
[1]Department of Psychology, Davidson College, Davidson NC, USA
[2]Language Evolution, Acquisition, and Development Group, Newcastle University, Newcastle upon Tyne, UK

Text generated by large language models (LLMs) is now often indistinguishable from text generated by humans. Designers of these models claim they are perilously close to reaching human-like levels of general intelligence (e.g. Bengio et al., 2023), with progress rapidly advancing as models become "multimodal" (though note that this is confined to the ability to integrate text and images, and does not approach the extent of multimodality in human language; Goldin-Meadow, 1999; Rasenberg et al, 2022). Alongside this, some cognitive scientists have declared that either impressive LLM performance (Contreras Kallens et al., 2023; Piantadosi, 2023; Frank, 2023) or its specific shortcomings (Katzir, 2023; Chomsky et al., 2023) provides compelling new evidence for longstanding debates about domain specificity and innateness of language (see Pleyer & Hartmann, 2018, for a summary of these debates in language evolution particularly).

The current work begins by questioning the immediate relevance of LLMs for understanding human language. We situate the language capacity of LLMs in a comparative perspective with the human language capacity using a Tinbergian framework of mechanisms, development, adaptive function and phylogeny. While LLMs share narrow adaptive function with human language (with coverage for producing text only, which is a proxy of only some spoken languages with written forms), the way in which an LLM develops its language capacity ("ontogeny") and the mechanisms by which it stores linguistic representations differ fundamentally. Despite the technically shared ancestry of human language and LLMs, we situate the phylogeny of LLMs as being a case of an *analogous* trait to human language, representing (if anything) partially overlapping convergent evolution, accomplished by means of human design rather than natural selection.

While this places the behaviour of LLMs as fundamentally different from human language in many important respects, it nonetheless provides us with one of the first close *functional* comparators for human language. The performance of LLMs alone cannot tell us much about how the human language faculty works, but careful probing of the difference in performance between humans and LLMs on specific, language evolution-oriented tasks provides an opportunity for new insights.

To this end, we report tests of a series of Artificial Language Learning (ALL) tasks focusing on training on partial systematic vocabularies with structured meaning spaces (adapted from Kirby, Cornish, Tamariz and Smith, 2014) followed by a test on full vocabulary. We contrast multimodal language models with text only models. Text only models were relatively successful in learning labels for seen items and recalling the meanings of novel words (unseen items). Some multi-modal models had comparable performance for learning labels, but when asked to generate descriptions of images for novel words results were inconsistent. A subset of multimodal models also showed uneven performance, often "collapsing" a systematic vocabulary by reproducing the same label repeatedly, even for unrelated shapes - an underspecification found in early iterated ALL without communication pressures (Kirby, Cornish & Smith, 2008). While the source of this disparity remains unclear, we suggest that the vision models generally used in multimodal LLMs are ill-adapted to dealing with the kinds of simple geometric images often deliberately chosen for ALL with humans. However, this same simplicity makes these images easy to describe systematically to text only LLMs.

While the reason why multimodal models fail to capture simple geometric structures in images may be obvious (it likely reflects the predominantly photographic input on which these multimodal models were trained), the implications of this are nonetheless interesting. Humans also have predominantly complex visual input, particularly during our evolutionary history, and yet simple geometric shapes are used in ALL studies with humans precisely because they make structure in meaning spaces readily apparent. We report on ongoing work using text to image generation models to create structured photographic image sets for systematically testing ALL across multimodal models and humans.

References

Bengio, Y., Russel, S., Musk, E., Wozniak, S., et al. (2023, March 22). Pause Giant AI Experiments: An Open Letter. Future of Life Institute. https://futureoflife.org/open-letter/pause-giant-ai-experiments/ Accessed 13 October, 2023.

Chomsky, N., Roberts, I., & Watumull, J. (2023, March 8). Opinion: The false promise of ChatGPT. *New York Times*.

Contreras Kallens, P., Kristensen-McLachlan, R. D., & Christiansen, M. H. (2023). Large Language Models Demonstrate the Potential of Statistical Learning in Language. *Cognitive Science*, *47*(3), e13256. https://doi.org/10.1111/cogs.13256

Frank, M. (2023). Bridging the data gap between children and large language models. *Trends in Cognitive Sciences*. https://doi.org/10.1016/j.tics.2023.08.007

Goldin-Meadow, S. (1999). The role of gesture in communication and thinking. *Trends in Cognitive Sciences*, *3*(11), 419-429.

Hurford, J. R. (2011). *The origins of grammar: Language in the light of evolution II*. OUP Oxford.

Katzir, R. (2023). Why large language models are poor theories of human linguistic cognition. A response to Piantadosi (2023). Lingbuzz preprint. https://lingbuzz.net/lingbuzz/007190

Piantadosi, S. (2023) Modern language models refute Chomsky's approach to language. Lingbuzz preprint. https://lingbuzz.net/lingbuzz/007180

Pleyer, M., & Hartmann, S. (2019). Constructing a Consensus on Language Evolution? Convergences and Differences Between Biolinguistic and Usage-Based Approaches. *Frontiers in Psychology*, *10*. https://www.frontiersin.org/articles/10.3389/fpsyg.2019.02537

Rasenberg, M., Pouw, W., Özyürek, A., & Dingemanse, M. (2022). The multimodal nature of communicative efficiency in social interaction. *Scientific Reports*, *12*(1), 19111.