Imitation or Communication? Examining Language Models with Cooperative Context and Language Development

Zhiyuan Zhang
Department of Automation
Tsinghua University
z-zy20@mails.tsinghua.edu.cn

Abstract

With the rapid development of natural language processing, current researchers transited from previous task-specific models to the more versatile, unified capabilities exhibited by GPT-like models. While these models achieve fluency in natural language and excels at coding, this paper argues that a crucial aspect remains underexplored: the process of language learning and development in these models. We propose a novel framework to evaluate language models, not just on linguistic proficiency but on their ability to transfer communication knowledge to new languages rapidly like humans, under cooperative contexts. Central to this framework is the idea that genuine communication in language models should be grounded in tasks that require mutual help and involve modalities beyond the language used, and these models should display the ability to rapidly learn and develop new languages, akin to human capabilities. This is explored through a referential game and evaluated for different stages that mirrors human behaviour, from symbol/icon grounding to symbol simplification and, finally, the emergence of systematicity. This framework aims to shift the current paradigm in language model evaluation, emphasizing the importance of interactive, adaptive, and contextually aware communication abilities.

1 Introduction

Language models have traditionally served specific purposes, specially designed and trained for tasks such as sentiment analysis or language translation. With the emergence of GPT-like models, however, a remarkable shift occurred. GPT models, unified various language task representation and displayed great performance on various tasks[4]. However, as we marvel at these accomplishments, a fundamental aspect often goes unexamined: the process of transferring communication knowledge. In this essay, we explore the core foundation of language development – cooperative context – and propose a framework for language models that mirrors human language learning development. We argue that to establish genuine communication, a language model must not only be proficient in existing languages but also exhibit a remarkable ability to rapidly learn and develop new languages, akin to human capabilities.

2 Cooperative context: foundation for communication

In this section, we give some foundation assumptions or prior attributes before we develop or test language models' communication abilities.

2.1 Models: task driven

In human communication, the motivation often lies in task completion that required collaborative efforts[2]. Similarly, language models should engage in tasks that necessitate human collaboration or insights, differentiating true communication from mere mimicking. Current models, like GPT-4, are programmed for friendly responses, yet their interaction is largely limited to their training data in instruction tuning. A more effective approach would be involving these models in tasks that require active communication with humans or other models, thereby challenging them to apply their learned language in practical, real-world scenarios.

2.2 Task: another modality

To differentiate mimicking or true communication, the "language" used (vector tokens, English words, or even emojis) should be grounded with another modality. If constrained within the language used only, there exists no proof that the model is simply memorizing the training set. This another modality could be image, audio, or even a python interpreter. Only when the communication is actually grounded with another modality, it can be considered transferring information useful for cooperative problem solving.

2.3 Prior: mutual help

Humans communication are effective and efficient, and the reasons include that people communicate informative messages based on common conceptual grounds to cooperate[3][2]. Language models should also display this intention, or trained to serve this purpose.

3 Language development: proof of capability

Here we argue that, since humans can develop new communication tools rapidly under specific constraints (like symbols developed in referential games[3], or developing sign languages among groups of people with different native sign languages) , a language model, if really possessed the power of communication, should display similar rapid learning or development of new languages. We aim to test these learning ability, using setups from a referential game. Assume that for this game, a language model and a human are given a fixed set of vocabularies (say, 100 emojis or English words) and asked to play, and we expect the models with communication powers to develop efficient and effective "language" for communication and models that simply mimic language usage cannot. Here we segment this process into stages that mirror human language development, with each stage representing a greater challenge for AI models in terms of adaptive communication skills.

3.1 Stage 1: symbol / icon grounding

Humans' language starts with grounding objects or concepts to atomic words or characters[1]. For example, Chinese characters started as iconic figures of natural objects: circle for sun, lines for water courses. The iconic characters were invented and linked to their referees because they are similar in shape. This initial phase of communication establishes foundational links between different modalities. Or, these links may be established arbitrarily if no obvious resemblance exists (Saussure believed that natural language features the arbitrary link between sign and meaning but other researchers proposed non-arbitrary origins for a few English words [1]). Applying this to a referential game, humans or AI models should be quickly establishing signs or words that can ground the task description using the "language". For a language model playing the referential game, if we give them a fixed set of vocabulary of meaningless "words" like "bouba" or "kiki", models may establish arbitrary links; if given a fixed set of English words but not descriptive enough for the task, like an image set contains various animals but word set is for describing household items, we may expect the model to utilize a certain degree of iconicity, for example, maybe "table" for elephant images and "cup" for mouse images because of the similarity in sizes.

3.2 Stage 2: symbol simplification

As the communication goes on and it's happening under some time limit or encouragement to shorter sentence length, humans simplify symbols[1]. We expect models to behave alike: as time goes by

the model can simplify symbols while keeping the effectiveness of communication. Using the same example, in the beginning the model may use "Big table with four thick legs" for elephant images and "Tiny white cup with adorable handles" for mouse images, as the process goes on the model may use "table" for elephants and "cup" for mouses.

3.3 Stage 3: systematicity emergence

Finally we can expect some degree of systematicity to emerge in this language. For humans, there are evidence that systematicity exists in human natural language (e.g. the "three point water" component of a Chinese character indicate that the character is linked to water) and "languages" developed in games (e.g. in a You Draw, I guess game, when expressing "art gallery", "museum" and "theatre", participants started with special symbols for each one but then reached a level of systematicity where a fixed symbol on the left expressing a house, a "<" like symbol in the middle representing the relationship "in" and different symbols representing paintings, antiques and performers with curtains aside) [1]. During this process the atomic symbols are further fixed and reused, along with some structures or grammars to convey complex meanings. For the referential example, after establishing "table" for elephant images and "cup" for mouse images, we may expect the model to express white elephants as "white table" and brown mouses as "brown cup".

4 Conclusion

In this essay, we've discussed how we should test language models like GPT differently. Instead of just seeing how well they understand and generate language, we should also see if they can learn and communicate in new ways, just like humans do. We suggested using a referential game to test this, where models have to go through stages – starting with linking simple words to things, then making their language simpler, and finally, using language in a systematic way.

References

[1] Nicolas Fay, T. Mark Ellison, and Simon Garrod. Iconicity: From sign to system in human communication and language. *Pragmatics Cognition*, 22:243–262, 12 2014. doi: 10.1075/pc.22.2.05fay. 2, 3

[2] Steven Gross. Origins of human communication - by michael tomasello. *Mind Language*, 25: 237 – 246, 03 2010. doi: 10.1111/j.1468-0017.2009.01388.x. 2

[3] Angeliki Lazaridou, Alexander Peysakhovich, and Marco Baroni. Multi-agent cooperation and the emergence of (natural) language. *International Conference on Learning Representations,International Conference on Learning Representations*, Nov 2016. 2

[4] OpenAI OpenAI. Gpt-4 technical report. Mar 2023. 1