Evaluate Communication Capability of Language Model: A Brief Survey from Evolution View

Jingze Zhang
Department of Automation
Tsinghua University
jz-zhang21@mails.tsinghua.edu.cn

Abstract

Creating intelligent agents that genuinely comprehend human language has long been the aspiration of natural language processing scientists. However, most of the current evaluation metrics for language models fail to accurately distinguish whether a language model truly grasps the essence of communication or merely replicates human speech patterns. Furthermore, contemporary robots struggle to engage in non-verbal communication with humans. To address these two issues, this essay will first differentiate between the concepts of machines truly understanding communication and merely imitating it. Subsequently, we will analyze the origins of human language production and emergence within the context of evolutionary biology. Building upon this analysis, this essay will introduce two criteria for assessing whether intelligent agents have mastered the essential traits of human communication.

1   Introduction

Building intelligent agents or models capable of human-like communication has been a longstanding aspiration for scientists. Over the years, researchers have developed numerous language algorithms, language models, and multi-modal models that excel in various areas, such as language question-answering[2, 7], image captioning, video language queries, translation[6], and many-to-many-modals question and answering[10, 9, 3]. However, when talking about communication abilities, the current language models' are still far from humans.

In recent years, various metrics have been developed to assess the capabilities of language models[1]. However, these evaluation criteria primarily focus on semantic aspects, facing challenges such as being single-modal and lacking a comprehensive consideration of context and user background knowledge. To address these issues, this paper introduces two challenging tasks for evaluating language models' communication abilities. These abilities necessitate an intelligent agent to not only comprehend semantics effectively but also exhibit flexibility in grasping context, background knowledge, and common societal norms. Furthermore, they require the agent to have a more explicit communication intent and a heightened ability to infer the intentions of others.

2   Human communication ability: a evolutionary view

Communication is a milestone in human evolution. Based on deep understandings on context, backgrounds and intention, we can get information from others and cooperate with others smoothly. The evolutionary steps of human communication ability can be summarized into the following steps.

Early Communication   Early humans relied on basic forms of non-verbal communication, such as gestures, facial expressions, and body language, to convey information and intentions[5]. These forms of communication helped with hunting, gathering, and group coordination. Such simple communication can also be found in apes[5].

Verbal Communication   Building upon the foundation of early communication, humans gradually developed rudimentary languages[4, 5]. These basic languages were capable of conveying richer information, allowing individuals to gain a clearer understanding of each other's intentions. Subsequently, language continued to evolve and solidify, giving rise to the written word, through which beliefs, knowledge, culture, and emotions could be passed down within and across generations.

Building upon the analysis above, it's essential to mention that communication and language are two distinct concepts. Communication is the act of conveying intent and information, with language being just one means of achieving this. Communication appears not only in human societies but also in some lower-order social creatures, such as ants and bees. However, communication that involves the transmission of shared beliefs is unique to human communities. Therefore, pondering how to construct intelligent entities capable of communication is a crucial step toward achieving general artificial intelligence whose value is in aligned with human society.

3   Communication abilities evaluating tasks

The primary application scenarios for agents' communication skills are communication between humans and machines, as well as communication among intelligent agents[8]. Based on this, we have proposed two tasks to evaluate the abilities of agents in communication with humans and among intelligent agents.

3.1   Task 1: virtual AI bot test

The task design and inspiration is similar to the design of Turing test. Imagine your communicating process via Wechat or other social media. If the other agent behind the screen were not a real person but a language model, but you couldn't tell them apart, then we are convinced to say that the artificial agent has the communication ability similar to human.

Although the task sounds easy, it is actually an especially challenging one. For instance, large language models (LLM) such as chatGPT or GPT4 nowadays can be a good listener and provide excellent question answerer, however, their responses lack a high degree of diversity and are unable to engage in emotional interactions with the conversation partner. They resemble more of a question-and-answer machine rather than an intelligent agent capable of deep interactions, adept at listening, and skilled at asking questions when communicating with others.

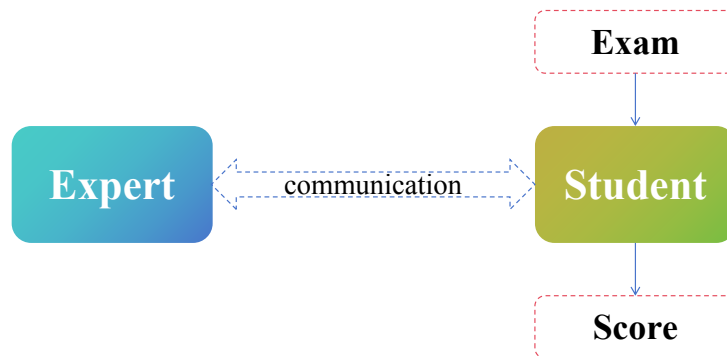## 3.2   Task 2: multi-agent cooperation

Figure 1: Pipeline of the *teacher-and-student-task*. The expert and student are 2 language model, where the expert comprehend knowledge in one professional field but the student knows nothing about the field. The expert can **only** teach the student agent via bidirectional verbal communication. After the teaching process, the student agent will complete an exam and the score can be served as a metric to evaluate the performance of the bidirectional verbal communication process.

Communication is an active process. For humans, our communication with others always occurs when we have a need to communicate. We continuously convey survival-related information, transfer knowledge, or engage in emotional interactions through communication. What's more communication enhances the smoothness of collaboration between individuals.

Based on the analysis above, we propose 2 simple but challenging evaluation sub-task to evaluate the communication abilities of agents in multi-agent scene.

We call the first sub-task *teacher-and-student-task*. The inspiration for this task stems from the significant aspect of human communication, which involves a substantial purpose of knowledge transfer. In the process of knowledge transmission, it is not just about one-way delivery from teachers to students but also the need for a closed-loop feedback mechanism where students can promptly ask questions to teachers, especially when encountering unclear points of knowledge. The pipeline of the sub-task is showcased in figure1.
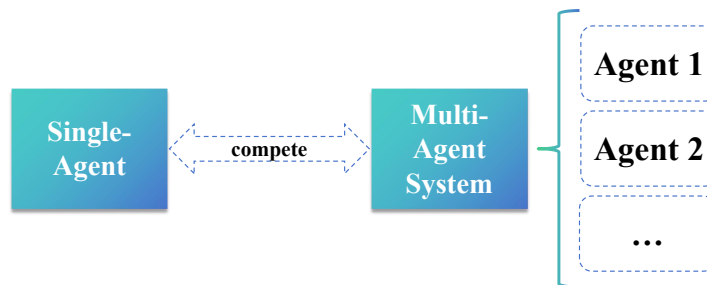

Figure 2: Cooperative-and-survival-task

We name the other sub-task *cooperative-and-survival-task*. The intuition of the sub-task comes from the initial function of human communication in the wild. The pipeline of the task is showcased in figure 2.

4 Summary

For intelligent agents, possessing the ability to communicate is a significant step toward achieving general artificial intelligence. It is through this skill that humans have distinguished themselves from the animal kingdom.

Drawing from our understanding of human communication, this essay introduces two tasks for evaluating whether intelligent agents have acquired the essential characters of human communication. If an agent merely mimic human communication without truly comprehending when and why to say or do something, they will be unable to fulfill these two tasks. Nevertheless, the path to genuinely teaching intelligent agents how to communicate is a long one. Before they can master the art of communication, they must first learn to understand intention, context, collaboration, and have the ability to watch and help.

References

[1] Stanley F Chen, Douglas Beeferman, and Roni Rosenfeld. Evaluation metrics for language models. 1998. 1

[2] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805*, 2018. 1

[3] Runpei Dong, Chunrui Han, Yuang Peng, Zekun Qi, Zheng Ge, Jinrong Yang, Liang Zhao, Jianjian Sun, Hongyu Zhou, Haoran Wei, et al. Dreamllm: Synergistic multimodal comprehension and creation. *arXiv preprint arXiv:2309.11499*, 2023. 1

[4] Nicolas Fay, Mark Ellison, and Simon Garrod. Iconicity: From sign to system in human communication and language. *Pragmatics & Cognition*, 22(2):244–263, 2014. 2

[5] Michael Tomasello. *Origins of human communication*. MIT press, 2010. 1, 2

[6] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017. 1

[7] Jason Wei, Maarten Bosma, Vincent Y Zhao, Kelvin Guu, Adams Wei Yu, Brian Lester, Nan Du, Andrew M Dai, and Quoc V Le. Finetuned language models are zero-shot learners. *arXiv preprint arXiv:2109.01652*, 2021. 1

[8] Michael Wooldridge. *An introduction to multiagent systems*. John wiley & sons, 2009. 2

[9] Shengqiong Wu, Hao Fei, Leigang Qu, Wei Ji, and Tat-Seng Chua. Next-gpt: Any-to-any multimodal llm. *arXiv preprint arXiv:2309.05519*, 2023. 1

[10] Zhengyuan Yang, Linjie Li, Kevin Lin, Jianfeng Wang, Chung-Ching Lin, Zicheng Liu, and Lijuan Wang. The dawn of lmms: Preliminary explorations with gpt-4v (ision). *arXiv preprint arXiv:2309.17421*, 9, 2023. 1