Paul Piwek

School of Computing and Communications

The Open University

Conversational large language models are speakers ?

+   −

Current generations of conversational LLM can pass time-limited versions of the Turing test (Jones & Bergen, 2023)

Conversational LLMs do not have the right Gricean (1957) intentions

−   −

The Turing test is not fit for purpose because of the Eliza effect (Weizenbaum, 1966)

Conversational LLMs haven't fully mastered **S**(peaking).

Can we be sure that training isn't providing machines with implicit intentions?

Reported intentions are typically after the fact rationalizations (Mercier & Sperber, 2017).

+

Conversational LLMs, such as ChatGPT and Gemini, apply different but equally rigid unengaged strategies when contradicted: Gemini (March, 2024) always cedes to the user ("You are correct. It appears …") whereas ChatGPT4o (July, 2024) never cedes ground ("To ensure absolute accuracy, I will recompute once again").

Definitions

S = speaking as a contribution by a person to a language game, i.e. a normative social activity requiring (i) sensitivity to, i.e. caring about, peer assessment of one's contributions and (ii) engagement with peer assessment of others' contributions.

A = the algorithmic generation of output strings that we take to be English or French or Chinese or . . . , given a (more or less formal) specification of requirements on the output (e.g. a prompt, logic formula or other) .

The chatbot conceit = the design of systems that do **A** but appear to be in the business of doing **S** by framing interactions as dialogue.

The pragmatics challenge = What are the ingredients **I** such that **A + I = S**?

References

Herbert Paul Grice. 1957. Meaning. Philosophical Review, 66(3):377–388.

Cameron Jones and Benjamin Bergen. 2023. Does GPT-4 Pass the Turing Test? arXiv preprint ArXiv:2310.20216 [cs]

Hugo Mercier and Dan Sperber. 2017. The Enigma of Reason. Harvard University Press.

Joseph Weizenbaum. 1966. Eliza a computer program for the study of natural language communication between man and machine. Commun. ACM, 9(1):36–45.