New Form of Intelligence?

- When applying tests designed for humans to LLMs, interpreting the results can rely on assumptions about human cognition that may not be true at all for these models => anthropocentric bias [1, 3]
  - => Opinion: intelligence and understanding “are the wrong categories” for talking about LLMs [1]
  - Open question: would it make sense to see the systems’ behavior not as “competence without comprehension” but as a new, non-human form of understanding? [1]
- Intelligent thought could be a mosaic of simple operations that, when studied up close, disappeared into its mechanical parts (c) Max Newman [2]
- „Real“ language understanding and intentionality consist of attributions of **unobservable** mental states [7] and it is unclear how we can meaningfully test for the „realness” of thoughts, feelings etc. [2]

=> Opinion: we need better measures for evaluating thought competence in LLMs before we can draw conclusions [7]

References

[1] The Debate Over Understanding in AI’s Large Language Models, Santa Fe Institute

[2] Do Large Language Models Understand Us?, Google Research

[3] Sparks of Artificial General Intelligence: Early experiments with GPT-4, Microsoft Research

[4] Meaning without reference in large language models, UC Berkeley & DeepMind

[5] On the Dangers of Stochastic Parrots: Can Language Models Be Too Big? 🦜, University of Washington et al.

[6] Dissociating language and thought in large language models, The University of Texas at Austin et al.

[7] Large Language Models: The Need for Nuance in Current Debates and a Pragmatic Perspective on Understanding, Leiden Institute of Advanced Computer Science & Leiden University Medical Centre