GPT和LLaMA2哪个“情商”更高

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/27337.html

本文仅供学习交流之用，版权归原作者所有，请勿用于商业用途！

GPT和LLaMA2哪个“情商”更高。《自然-人类行为》5月20日发表的一篇论文显示，在测试追踪他人心理状态能力——也称心智理论——的任务中，两类大语言模型（LLM）在特定情况下的表现与人类相似，甚至更好。

心智理论对人类社交互动十分重要，是人类沟通交流和产生共鸣的关键。之前的研究表明，像大语言模型这类人工智能可以解决复杂的认知任务，如多选决策。不过，研究人员一直不清楚LLM在心智理论任务——被认为是人类独有的能力——中的表现是否也能比肩人类。

在这项新研究中，德国汉堡-埃彭多夫大学医学中心的James Strachan和同事选择了能测试心理理论不同方面的任务，包括发现错误想法、理解间接言语以及识别失礼。

研究人员随后比较了1907人与两个热门LLM家族（由美国OpenAI公司开发的GPT模型和由美国Meta公司开发的LLaMA2模型）完成任务的能力。

他们发现，GPT模型在识别间接要求、错误想法和误导表现等方面能达到、有时甚至超越人类平均水平，而LLaMA2的表现则逊于人类水平。在识别失礼方面，LLaMA2强于人类但GPT表现不佳。

Strachan和同事指出，LLaMA2的成功被证明是因为回答的偏见程度较低而不是因为真的对失礼敏感，而GPT看起来的失利其实是因为对结论的保守态度而不是推理错误。

研究人员提醒道，LLM在心智理论任务上的表现堪比人类不等于它们具有人类般的能力，也不意味着它们能掌握心智理论。但他们也指出，这些结果是未来研究的重要基础，并建议进一步研究LLM在心理推断上的表现会如何影响个体在人机交互中的认知。（来源：中国科学报赵熙熙）

相关论文信息：https://doi.org/10.1038/s41562-024-01882-z

作者：James Strachan 来源：《自然—人类行为》

更多 科学进展 请访问 https://www.iikx.com/news/progress/

本文版权归原作者所有，请勿用于商业用途，爱科学iikx.com转发