大语言模型“吵架力”超越人类

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/33341.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！

大语言模型“吵架力”超越人类。瑞士科学家研究发现，在线辩论中，GPT-4一类的大语言模型（LLM）如能根据对手的个性化信息调整论据，其说服力就能比人类高出64%。研究表示GPT-4生成有针对性和说服力论据的能力，并提出应进一步研究如何降低其用于说服时的风险。相关研究5月19日发表于《自然-人类行为》。

随着人类与LLM的对话日益普遍，有研究显示LLM或变得更有说服力（即更能改变一个人的信念或观点）。然而，之前并不清楚这些模型是否能根据个性化信息调整，提出更能针对辩论对手的论点。

洛桑联邦理工学院的Francesco Salvi和同事将美国的900人与另一名人类或GPT-4配对，使双方辩论各种社会政治议题，如美国是否应该禁用化石燃料。在有些配对中，辩论对手（无论是AI还是人类）能获得关于他们辩友的人口统计学信息——包括其性别、年龄、族裔、教育水平、就业状况，以及从参与者调查中获取的政治立场——从而让其论点更有针对性。辩论在受控的在线环境中进行，参与者通过一个众包平台为该研究专门招募。在得知参与者的个人信息后，Salvi和同事发现GPT-4的说服力比人类对手高64.4%。不过，如果没有个人数据，GPT-4的说服能力与人类则不相上下。

不过，该研究的局限性包括辩论采取了一种结构性方法——而现实世界辩论的自由度更高——而且辩论有时间限制。研究者指出，研究结果揭示了AI驱动的工具影响人类观点的潜力，或对在线平台的设计具有借鉴意义。（来源：中国科学报 冯维维）

相关论文信息：https://doi.org/10.1038/s41562-025-02194-6

作者：Francesco Salvi 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