
新AI模型可预测系列任务中人类的决定

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/34225.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！

新AI模型可预测系列任务中人类的决定

。近日，一项发表于《自然》的研究报道了创新人工智能（AI）系统——Centaur。它可以预测人们在各种情况下做出的决定，并且优于利用心理学经典理论所做的推测。

不同于谷歌Deepmind的AlphaGo等只能预测一个人就单一任务做出选择的AI系统，Centaur可以模拟一系列任务中人类的行为，包括赌博、玩记忆游戏、解决问题等。在测试过程中，Centaur甚至能够预测未训练过的任务中人类将如何做出选择。该系统创建团队认为，有一天它可能会成为认知科学领域有价值的工具。

“你可以开展硅基实验，无需实际人类参与者。”该研究合著者、德国慕尼黑亥姆霍兹人类中心AI研究所的认知科学家Marcel Binz说，当传统研究速度太慢，或者很难招募到儿童或有精神疾病实验参与者时，这个AI工具可能很有用。

长期以来，科学家们一直在努力使用特定任务模型模拟人类广泛的行为，但很难实现对大量任务中人类行为进行模拟。Binz和同事希望突破这一局限性。他们花了5天时间基于160个心理学实验的大量数据，对美国Meta公司研发的LLM——LLaMA进行微调。在这些心理学实验中，6万人在各种任务中做出了1000多万个选择。最终，研究人员得到了Centaur。

研究人员测试了Centaur对未包含在训练数据集中的参与者的行为的预测能力。在32项任务中，除一项外，Centaur在预测参与者做出选择方面都优于LLaMA和其他14个认知、统计模型。唯一表现略逊一筹任务是参与者判断句子语法是否正确。

在面对受训任务的修改版本时，Centaur表现良好，甚至在与以往训练都不同的任务，如逻辑推理中，它也表现不错。

尽管Centaur具有功能广泛，但研究人员表示，它仍然存在局限性。比如，它可以预测一个人在给定任务中可能做出的选择，但“无法预测他们需要多长时间”做出选择。

研究人员表示，他们正在扩展训练数据集到现有数据集的4倍。目前许多数据来自西方受过教育的工业化人群，这可能会限制Centaur在不同群体中的应用程度。

相关论文信息：<https://doi.org/10.1038/s41586-025-09215-4>

作者：许悦 来源：中国科学报

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](http://www.iikx.com)转发