
AI可测试性格，并预测人类反应

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/41749.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！

AI可测试性格，并预测人类反应。以色列科学家开发出一种利用ChatGPT从任意文本中生成性格评估问卷的方法，他们将其应用于临床医生用来诊断精神障碍的著名文本DSM-5，并作为一个非常规例子，将其应用于一本占星术教科书。他们发现ChatGPT不仅可以用来创建和验证这些问卷，而且还可以在进行调查之前准确地预测人们的反应。8月6日，相关研究发表于交叉科学期刊*science*。

ChatGPT和其他可公开访问的大语言模型（LLMs）是在互联网上通过编译来自网站和社交媒体的数万亿人类语言数据进行训练的。基于这种训练，研究人员一直在考虑LLMs是否对人类语言有专家级的理解，并由此扩展到它们的算法中。

考虑到性格特征反映在语言中，LLMs可能已经学习了人类性格的结构，这是训练它们的自然副产品。来自耶路撒冷希伯来大学的主要作者Rotem Monsa说，所以，虽然它们没有专门被教授心理学或人格理论，但这些理论已经融入了LLMs学习的语言中。

为了测试LLMs自然评估人类性格的能力，研究人员使用GPT-4生成了两份性格评估问卷。他们首先使用了DSM-5的摘录作为源文本，并假设人格特征是在人格障碍连续谱上定位的。作为对照，另一份调查问卷使用了占星学教科书。对于前者，ChatGPT根据DSM-5中的人格障碍描述生成了一份带有人格陈述的问卷。例如，根据偏执型人格障碍部分，问卷要求参与者对经常怀疑他人的动机或发现很容易信任别人等陈述的认同程度进行排名（1=非常不同意，直到5=非常同意）。占星术问卷中也出现了类似的表述，但这些是基于源文本中对占星学黄道十二宫所赋予的人格特质进行的分析。

Monsa说：我们希望选择的文本能够非常详细地描述人类的个性，但同时又在科学依据方面呈现出截然不同的特点。DSM-5经过数十年的临床研究而不断得到完善，是临床精神病学的标准诊断手册，享誉全球。而这份占星学文本则非常注重文化背景，但并不具备科学验证的依据。

基于LLMs的问卷生成后，这些问卷被分发给600名参与者，同时附上五大性格特质问卷（BFI），这是目前验证程度最高的性格问卷，以评估其实用性。

基于DSM-5的问卷结果显示，人格集群内部具有较高的内在一致性；这意味着在现实世界中通常相关的特征（如回避和依赖）在参与者的回答中也相关。重要的是，这些结果与BFI的结果相一致，这一发现有助于验证问卷在衡量人类心理的真实生活模式方面的力量。正如预测的那样，相比之下，占星术问卷显示出了较弱的内部一致性。我们的数据表明，占星因素并不能反映连贯的心理层面。Monsa说，例如，在真实人群中，火相元素的性格特征实际上并不相同。

尽管存在这种限制，但两份调查问卷均能够通过受访者的回答来预测诸如抑郁、焦虑和幸福感等生活方面的结果，其预测水平可与BFI相媲美。这表明，尽管将性格特征分配给星座没有科学依据，但LLMs从这些文本中提取的与性格相关的内容仍保留了有意义的心理信号。

然而，最令人惊讶的结果是，ChatGPT可以预测参与者在接受问卷调查之前的反应。对于这两份问卷，ChatGPT预测的平均回答和问题之间的相关性具有很高的现实世界的准确性，这表明LLMs在人口水平上对人格动态有一种天生的理解。

LLMs在看到任何人类数据之前就可以预测人类的反应模式，这一事实表明，这些模型已经观察到一些关于人类心理的普遍有意义的东西。Monsa说，这些结果告诉我们，LLMs不仅是一个很好的内容生成工具，而且还可以作为自己产出的知情评估者。

然而，作者指出，如果这些方法在不同的语言和文化中重复使用，结果可能就站不住脚了。我们会认为，在其他语言和文化中，结果不会像我们在这里看到的那样强烈，因为LLMs主要是在西方文化中接受英语文本的培训，Monsa说，我们认为这真的很有趣，我们实验室的几名成员目前正在使用其他语言的LLMs进行测试。

综上所述，这些结果表明LLMs可以帮助研究人员从任何来源材料中快速创建基于人格的问卷，并直接测试其临床使用的有效性。此外，临床和心理学见解与基于人工智能的工具的力量相结合，是临床和实验心理学发展的一个变革阶段，可能会对该领域产生深远影响。（来源：中国科学报 冯维维）

相关论文信息：<http://dx.doi.org/10.1016/j.isci.2026.116909>

作者：Rotem Monsa 来源：《交叉科学》

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](https://www.iikx.com)转发