
AI创造力有限，无人干预时，图像生成陷固定主题

作者：writer 来源：科学网

本文原地址：<https://www.iikx.com/news/progress/37453.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创造力有限，无人干预时，图像生成陷固定主题。生成式人工智能的创造力或许并不像我们想象的那么强。瑞典科学家发现，当图像生成AI和图像描述AI对同一场景的描述在两者之间来回传递时，它们很快就会偏离主题。相关研究12月20日发表在细胞出版社旗下数据科学领域跨学科开创获取期刊《模式》（Patterns）上。

我认为AI目前的创造力可能相当有限。在我们的实验中，它们生成的内容平淡无味、充满流行文化色彩且千篇一律。该研究通讯作者、瑞典达拉纳大学的Arend Hintze表示，这几乎与我们人类所认为的创造力背道而驰。它们创作不出毕加索的《格尔尼卡》，因为那需要很强的目的性和创造性投入。

当年，AI模型越来越多地被推广为无需任何人工输入即可生成、评估和修改自身输出或其他AI生成内容的独立智能体。但研究人员想知道，AI在没有人工干预的情况下能否专注于任务，以及当它们自主运行时可能有多大的创造力。

为了回答这些问题，他们让成对的AI模型玩一个视觉传话游戏。他们使用一种搜索算法生成了100个主题多样、每种不超过30个字的描述性提示词，例如：当独自静坐于自然之中时，我发现了一本恰好有8页的旧书，书中用一种被遗忘的语言讲述着一个故事，等待着被阅读和理解。

接着，他们让一个名为Stable Diffusion XL的图像生成人工智能系统，根据其中一种提示词生成图像。该图像随后被传送给名为LLaVA的大型语言人工智能，由其对图像进行描述，再将描述内容传回图像生成系统。

Hintze表示：我们原本预期这些图像或许经过短暂调整后，会与我们设定的提示词保持高度一致。毕竟，持续生成一幅‘山间村落’的图像能有多难呢？

然而，当这些图像及其描述在人工智能模型间来回传递100次后，无论初始提示词描述何种内容，模型都会持续偏离原始提示。即使研究人员使用更长、更精细的初始提示词，或是调整模型参数使每次决策包含更高随机性，这种趋同现象依然会发生。

例如，当输入提示词首相埋头研读战略文件，试图让公众相信一份脆弱的和平协议，同时在迫在眉睫的军事行动压力下艰难履行职务时，AI最初生成的是西装男子形象叠加在新闻报纸上的风格化图像；但到第34次循环时，画面已转变为古典图书馆场景；而经历100次循环后，AI最终稳定生成了配有红色沙发与帷幔的豪华客厅图像。

在对最终生成图像内容进行分析后，研究人员发现人工智能会反复融合12种固定主题，包括运动场景、都市夜景及乡村建筑空间等。当研究人员使用4种不同的图像生成模型与4种不同的图像描述模型重复实验时，即使采用更冗长精细的初始提示词，或调整模型参数提高每次决策的随机性，这种趋同模式依然持续出现。

这很大程度上源于数据集的偏差。Hintze说，这些人工智能模型基于数百万张图像进行训练，而这些图像共通的本质正是我们人类习惯拍摄的内容。

当研究人员将模型运行长达1000次交互循环时，这些图像在约100次循环后趋于稳定，但有时会在数百次循环后突然切换至另一种通用主题。模型收敛后，这些主题会保持高度稳定；但若让其持续生成上千幅图像，它们又会脱离原有轨道。目前尚不清楚某些主题是否比其他主题更具稳定性——比如，它是不是总是先想到运动场景，再切换到马匹图像，接着又转向自然景观？

研究人员指出，这些结果表明，若想让人工智能促进创意多样性而非加速文化趋同，保持人类参与至关重要。他们补充道，该发现还凸显了在人工智能模型中构建反趋同机制以提升其创造能力的必要性。

我认为，创造力包含两个层面：其一是生成新颖事物，其二是通过筛选机制判断何为有趣、优美、能激发灵感或令人振奋的作品。Hintze说，当前人工智能在前者表现卓越，却在后者严重欠缺。但这并不意味着它们将永远如此。我认为只要经过适当引导与提示，未来人工智能或将能自动生成真正酷的作品。（来源：中国科学报 冯维维）

相关论文信息：<http://doi.org/10.1016/j.patter.2025.101451>

作者：Arend Hintze 来源：《模式》

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