新工具可识别AI“代笔”

作者：writer 来源：科学网

本文原地址：https://www.iikx.com/news/progress/23426.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！

新工具可识别AI“代笔”。

人工智能(AI)聊天机器人ChatGPT的首次亮相引起了全世界的热议，因为它具有强大的文本处理和对话能力。不过，许多线索可以帮助人们区分机器人和人类。

美国科学家开发了一种工具，可以识别人工智能生成的学术文本，准确率超过99%。相关研究近日发表于《细胞报告物理学》。

我们努力创造一种易于使用的方法，解决辨别人工智能写作问题。如此一来，高中生也可以建立一个针对文本的人工智能探测器。论文第一作者、堪萨斯大学教授Heather Desaire说。

目前，人工智能写作存在一些明显的问题。Desaire说，最大的问题之一是，它汇集了许多来源的文本，却没有任何形式的准确性检查。

尽管许多人工智能文本检测器在网上可用，并且表现相当好，但它们并不是专门为学术写作而构建的。为了填补这一空白，该团队希望构建一个具有更好性能的学术写作检测工具。他们关注的是观点文章——科学家撰写的特定研究主题的概述。该团队选择了64个主题，并在相同的研究主题上创建了128篇ChatGPT生成的文章来训练模型。当他们比较这些文章时，他们发现了人工智能写作的一个指标——可预测性。

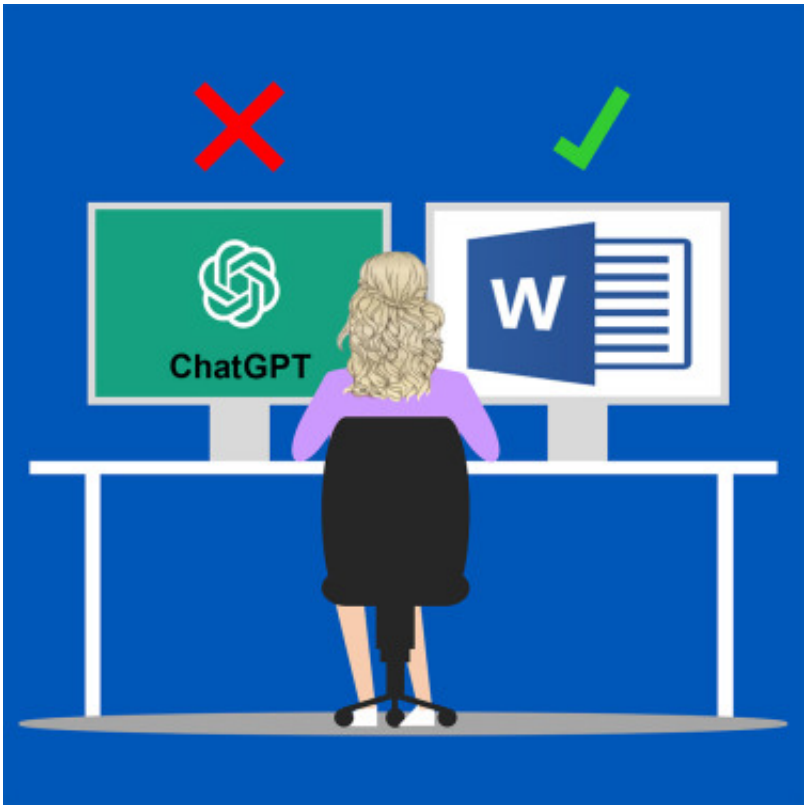
与人工智能相反，人类写作的段落结构更复杂，每段的句子数量和总字数都不一样，句子长度也不稳定。此外，对标点符号和词汇的偏好也是一个线索。例如，科学家倾向于使用然而、但是和尽管这样的词，而ChatGPT在写作中经常使用其他人和研究人员。最终，该团队为新模型列出了20个指标。

经过测试，对于整篇观点文章，该模型在区分人类和人工智能作者的准确率达到了100%。对于识别单个段落，该模型的准确率为92%。该模型远远超过了市场上现有的人工智能文本检测器。

接下来，研究人员计划确定模型的适用范围。他们想在更广泛的数据集和不同类型的学术写作中测试它。随着人工智能聊天机器人的不断发展，研究人员也想知道该模型能否跟得上形势。

当人们听到这项研究时，他们可能首先想到‘我可以用它来判断我的学生是否真的写了论文吗?’Desaire说。尽管该模型在区分人工智能和人类作者方面非常熟练，但Desaire表示，它的设计初衷并不是为了区分人工智能生成的学生论文，供教育工作者使用。不过，她指出，人们可以很容

易地复制他们的方法来针对自己的目的建立新模型。(来源：中国科学报 冯丽妃)



美国科学家设计的一种工具可识别ChatGPT代笔的蛛丝马迹。图片来自：Heather Desaire

相关论文信息：http://doi.org/10.1016/j.xcrp.2023.101426

作者：Heather Desaire 来源：《细胞报告物理学》

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转发