**快评丨ChatGPT写假新闻惊动警方，谣言该如何止于“智”者？**

[红星新闻](https://author.baidu.com/home?from=bjh_article&app_id=1561825967470087)2023-02-17 18:36

这一切，并不代表要把人工智能当成洪水猛兽，而是启发我们，对待人工智能创造性思维和批判性思维并存

我们常常说，谣言止于智者，现在好了，谣言起于“智”者——“智”是人工智能。

2月16日，一则杭州市政府3月1日取消机动车依尾号限行的“新闻稿”在网上疯传。这则新闻稿的来源是杭州某小区业主群讨论ChatGPT时，一位业主开玩笑说尝试用它写篇杭州取消限行的新闻稿，随后在群里直播了使用ChatGPT的写作过程，还把文章发在群里。其他业主不明就里，截图转发，导致错误信息传播。据报道，经杭州警方证实，该消息不实，警方已介入调查。



有人说，这则假新闻及其引发的后续效应，可以载入人工智能发展史了。虽然只是一个小小的涟漪，整个杜撰过程也是在小区群这个半公开的场域中进行的，当事人及时承认错误让事件并未造成实质性的破坏，但是这件事，确实第一次印证了大家关于人工智能的“超能力”是否会被用来作恶的联想。

因为作为一种深度学习型人工智能技术，ChatGPT可以说是“天生造假者”。它已经在各行各业展示出用武之地，写代码、译文献、写小说、写文案、编菜谱、写情诗……大家的探索热情，不断开发它的可能性，同时也对其进行着训练。就像任何工具都有两面性，ChatGPT也不例外，正因为它太好用了，学生用来代写论文，求职者用它编造简历，好事之徒用它来写假新闻，这些负面影响已经显现。而基于ChatGPT的“好用性”，这些假作品，几可乱真。有人据此提出了人工智能程序的“武器化”风险。

并非危言耸听，今年1月，新闻可信度评估机构News Guard的分析师指示ChatGPT对虚假信息作出回应，ChatGPT为100条虚假信息中的80条制作了更具细节的新闻报道、文章和电视脚本。对于那些不熟悉样本信息所涉内容或主题的人来说，ChatGPT制作的这些内容很容易被看作是合法、权威的信息。News Guard联合执行官格洛维茨称：“ChatGPT可能成为互联网有史以来最强大的传播虚假信息的工具。”

一些实验表明，ChatGPT等个性化的实时聊天机器人在制造阴谋论方面甚至相比人类阴谋论者更具优势，它能够修正糟糕的语法或错误的翻译，并变换表达方式，使阴谋论显得更具说服力和可信度。

比如这次取消限行的假新闻中，ChatGPT编的稿子有鼻子有眼，措辞、结构、语态跟一般的新闻通告非常像，连身边的一些媒体老编辑都信以为真。“此次取消限行是市政府为方便市民出行、优化城市交通而作出的重要举措”这种“官方表态专业用语”，可谓一本正经地胡说八道。真应了那句话——流氓不可怕，就怕流氓有文化。

本身ChatgGPT的一个大缺陷就是非常不擅长事实核查，在此基础上获得造假的指令后，内容生产的造假效率便会大大提高，造假成本简直低到尘埃。防止人工智能生产的虚假信息涌入这一问题，已经“怼”到了大众面前。

该怎么做不难想象。首先，人工智能缔造者有义务，从技术修正的角度，识别和排除可能会造成风险的造假指令，或者防止人工智能生产的内容，进入某些特定领域。其次，构建人工智能应用的伦理和法律框架，要和技术研发齐头并进。**如果缺少伦理和法律这两条“缰绳”，那么人工智能的野马“狂飙”起来后果难以预料。**第三，用户教育要同步进行。大众对人工智能生产内容（AIGC）新纪元感知是滞后的，并未意识到新技术给信息环境造成的巨大影响，对信源真实性的警觉尚未建立，对不同媒介的内容质量也没有基本认知。

这一切，并不代表要把人工智能当成洪水猛兽，而是启发我们，对待人工智能创造性思维和批判性思维并存。就像ChatGPT在接受《时代》专访时回答的：我还有很多局限，但人类应准备好应对AI。

红星新闻特约评论员 罕山

编辑 赵瑜

红星评论投稿邮箱：hxpl2020@qq.com