

每经记者：文巧 每经编辑：兰素英

4月6日，一则消息在推特上炸开了花。据路透社报道，澳大利亚墨尔本西部赫本郡的市长布赖恩·胡德（Brian Hood）指控OpenAI旗下的ChatGPT对其进行诽谤，或将对该公司提起诉讼，因为该聊天机器人在回答问题时错误地声称他是贿赂丑闻的有罪方。

值得注意的是，一旦正式提起，这将是全球首例针对生成式AI的诽谤诉讼。前加拿大互联网政策与公共利益诊所政策顾问、现香港中文大学法律系副教授Stuart Hargreaves在推特上评论道，“生成式AI和诽谤索赔的交集将很快成为一个非常有趣的法律问题。”

实际上，ChatGPT生成的虚假信息带来的担忧不是首次，这种担忧也正在引起广泛关注。据《华盛顿邮报》5日报道，一位名叫乔纳森·特利（Jonathan Turley）的法学教授也被ChatGPT错误地描述为曾对他人进行性骚扰，甚至还引用了不存在的报道作为信息来源。

随着生成式AI造成的虚假信息泛滥，ChatGPT等工具受到诽谤诉讼或许只是时间问题。而这类事件也向世界提出了一个复杂的问题——谁该对AI的虚假信息传播负责？

如今看来，AI的发展已超乎人类想象，监管机构可能还并未准备好如何解决这个问题。研究人员认为，政府应该扮演更多的角色，包括对模型的数据收集施加限制、对AI硬件的使用实施监控等。而AI开发人员、社交媒体平台和政府机构之间还应该开放新的合作方式，来帮助AI避免发布误导性信息。

图片来源：视觉中国-VCG111421152864

澳洲市长被诽谤，或索赔超90万

据报道，胡德此前曾在一家名为Note Printing Australia的公司工作，当时，他向澳大利亚监管机构举报了公司内部向外国官员行贿以赢得货币印刷合同的情况。胡德在接受媒体采访时表示，他不仅揭露了Note Printing Australia的行贿行为，还“成为了控方证人，并经历了无数法庭案件的所有过程”。

但在ChatGPT的描述中，胡德“被指控为在1999年至2005年期间贿赂马来西亚、

印度尼西亚和越南的官员，他在承认两项虚假指控后被判处30个月监禁。”

“我从未被指控任何事情。”胡德说道。“真正令人不安的是，其中有些段落却是绝对精确的，包括数字、姓名、日期、地点等。

这件事发生在今年3月。据悉，胡德已于3月21日向开发ChatGPT的公司OpenAI提交了一封“担忧信”。根据澳大利亚的法律，受害人必须发出通知解释情况，并要求OpenAI在28天内采取某种方式进行纠正。不过，截至目前，胡德方面还未收到OpenAI的回复。

胡德方面表示，如果OpenAI不纠正ChatGPT关于其曾因贿赂入狱服刑的虚假说法，他可能会起

诉OpenAI，这将是针对生成式AI的第一起诽谤诉讼。负责该案件的Gordon Legal律师事务所合伙人James Naughton认为，“从某种意义上说，这可能是一个具有里程碑意义的时刻，因为它将诽谤法应用于AI和IT新领域。”

澳大利亚的诽谤赔偿金通常上限为40万澳元左右（约合人民币184万元）。不过，Naughton表示，胡德目前不清楚读取到有关其虚假信息的确切人数，而这是赔付金额的决定因素。但诽谤性言论的性质非常严重，他可能会索赔超过20万澳元（约合人民币92.1万元）。

Naughton表示，一旦胡德提起诉讼，指控的方向可能是ChatGPT的回答中没有包含脚注，这将给用户一种错觉——它是准确的。“要追究‘算法是如何得出这个答案的’非常困难的，（ChatGPT）非常不透明。”他说道。

实际上，ChatGPT虚假信息带来的担忧不是首次，这种担忧也正在引起更多主流媒体的关注。

据《华盛顿邮报》4月5日报道，一位名叫乔纳森·特利（Jonathan Turley）的法学教授收到了一封令人不安的电子邮件。邮件显示，作为研究的一部分，加利福尼亚州的一位律师要求ChatGPT生成一份曾对他人进行过性骚扰的法律学者名单，而特利的名字就在名单上，ChatGPT还引用了《华盛顿邮报》2018年3月的一篇文章作为信息来源。

然而，问题在于，《华盛顿邮报》并不存在这样的报道。而特利也表示，他从未被指控骚扰学生。“这真令人不寒而栗。”他说。“这种指控非常有害。”

假新闻泛滥，谁来负责？

外媒评论认为，随着虚假信息的泛滥，ChatGPT受到诽谤诉讼只是时间问题。像ChatGPT这样发布潜在诽谤内容的新生成式AI工具可能会带来下一个复杂的问题——谁该对AI的虚假信息传播负责？

然而，AI的快速发展已经超乎人类的想象，监管机构可能还并未准备好如何解决这个问题。“从法律的角度来看，我们不知道当就虚假信息起诉AI制造商时，法官会如何裁决。”法律专家杰夫·科塞夫（Jeff Kosseff）说。“以前没有经历过这样的事情。”

以美国为例，在互联网兴起之初，美国国会通过了一项名为第230条（Section 230）的法规，该法规保护在线服务免于对其托管的第三方创建的内容承担责任，例如网站上的评论者或社交应用程序的用户。但专家表示，尚不清楚若科技公司因自己的AI聊天机器人制作的内容被起诉，是否能够使用这种保护措施。

而以胡德案为例，悉尼大学媒体法教授大卫·罗尔夫（David Rolph）表示，以诽谤罪起诉OpenAI会很复杂，因为还需要考虑管辖权问题。

“我们与许多在线中介机构之间存在的一个问题是管辖权的基本问题，你真的可以在澳大利亚法院对他们提起诉讼吗？”他说道。“很多这些互联网中介机构都设在海外，其中很多在美国，这往往会引发各种各样的问题。”

在ChatGPT爆红之初，其凭借编写计算机代码、创作诗歌和交互能力引起极大关注，但当时已有许多专家提出警示，这种创造力也可能成为错误主张的引擎，AI可以借此极大地歪曲关键事实，甚至可以捏造主要来源来支持自己的主张。

为什么AI能够生成如此多的虚假信息？AI语言系统的好坏取决于它们接受训练的文本，但互联网通常充满了文化偏见、谎言和仇恨言论。

乔治敦大学安

全与应急技术中心的几位研

究人员在一篇论文中提到，

要遏制此类信息，政府应该扮演更多的角色，包括对模型的数据收集施加限制、对AI硬件的使用实施监控等。此外，AI开发人员、社交媒体平台和政府机构之间还应该开放新的合作方式，来帮助AI避免发布误导性信息。

目前来看，各地监管机构的确正在努力进行监管。

继意大利之后，当地时间4月4日，加拿大隐私专员办公室宣布对ChatGPT开发公司OpenAI展开调查，该调查涉及“OpenAI未经同意收集、使用和披露个人信息”的

指控。

每日经济新闻