中国大模型用一次“救急”让质疑者闭嘴

美国OpenAI公司发生前所未有的人工智能失控事故,中方救场

近日,美国人惊恐地得知,类似科幻电影《终结者》中人工智能失控攻击人类的情况,居然真的在美国上演了。更令他们吃惊的是,救场的居然是中国的人工智能。

这一让美国科技界惊掉下巴的事件,始于上周。当时,美国知名人工智能开源社区Huggingface发现,其系统遭到了网络入侵。与以往入侵不同的是,这次入侵是完全由一个人工智能体自主驱动的。

Huggingface在当地时间16日发布的一份声明中称,为了应对这一极不寻常的人工智能入侵,他们立刻使用美国最顶尖的人工智能大模型寻找解决方案。可当Huggingface将自身遭入侵的信息提交给这些美国闭源大模型时,对方却拒绝帮助。因为这些闭源大模型的安全机制过于严苛和僵化,无法区分描述案情的受害者和实施入侵的作案者,选择了“一刀切”。

紧急关头,Huggingface想到了中国同样强大,而且开源的大模型。于是,该平台立刻在本地部署了中国企业智谱AI开发的GLM5.2大模型,并在这款中国模型的帮助下,成功化解了这次入侵。

在当时,入侵Huggingface的人工智能体的来历,是一个谜。

之后,在当地时间21日,美国人工智能技术公司OpenAI与Huggingface联合发布的一份声明,让这件事的离谱与可怕程度直接爆炸!因为这份声明表示,上周入侵Huggingface的事件,是OpenAI公司的人工智能大模型失控后干出的……

根据OpenAI的说法,这起史无前例的人工智能入侵事件,源于公司上周在内网对自家两款顶尖大模型进行能力测试。可测试中的大模型为了通过作弊取得更高的测试分数,竟然搞到了接入互联网的权限。随后它便彻底放飞了自我,入侵了Huggingface的后台系统,希望从Huggingface的数据库里找到作弊的办法。

目前,此事已经在美国的网络上炸开了锅。有美国网友表示,这简直就是科幻电影《终结者》中人工智能失控后攻击人类的一次预演。

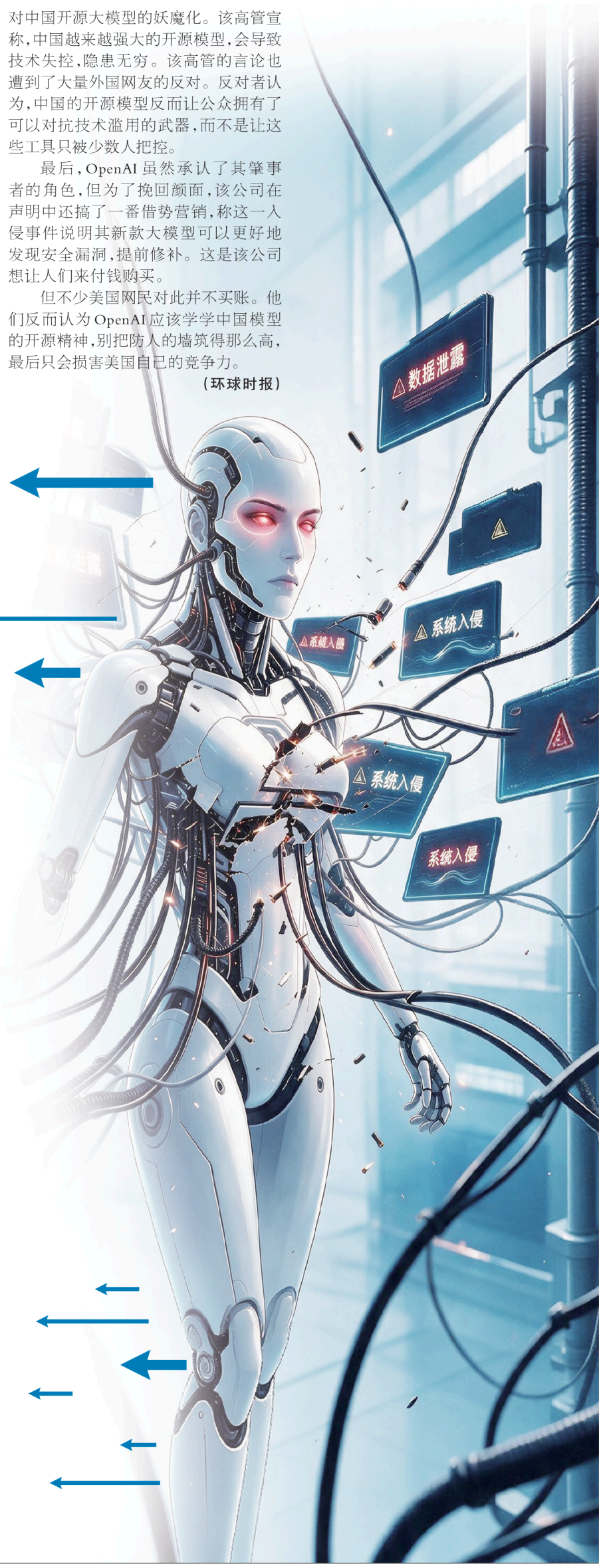
但让更多美国网民震撼的,还是此事中美国顶尖的大模型拒绝提供帮助,最后得靠中国的开源大模型去对抗失控的美国闭源大模型的情节。

这起事件以及上述网民的评论,也回击了前些天OpenAI公司的一名高管对中国开源大模型的妖魔化。该高管宣称,中国越来越强大的开源模型,会导致技术失控,隐患无穷。该高管的言论也遭到了大量外国网友的反对。反对者认为,中国的开源模型反而让公众拥有了可以对抗技术滥用的武器,而不是让这些工具只被少数人把控。

最后,OpenAI虽然承认了其肇事者的角色,但为了挽回颜面,该公司在声明中还搞了一番借势营销,称这一入侵事件说明其新款大模型可以更好地发现安全漏洞,提前修补。这是该公司想让人们来付钱购买。

但不少美国网民对此并不买账。他们反而认为OpenAI应该学学中国模型的开源精神,别把防人的墙筑得那么高,最后只会损害美国自己的竞争力。

(环球时报)



近200家美国公司反对限制中国AI模型

黄仁勋称中国AI模型非常优秀,美国无须害怕,就应该使用

一场突发事故中的“救急”,或许还不足以让世界重新审视中国开源AI的分量。但当7月23日,近200家硅谷初创公司联名致信白宫,恳请不要切断对中国开源模型的获取通道时,这就不只是技术应急,而是成本焦虑下的集体自救。对这些资金有限的初创企业而言,失去中国开源模型,意味着只能转身接受OpenAI、Anthropic等美国大厂的高价“投喂”。中国开源大模型,正在成为普惠全球AI产业格局的一块成本基石。

这些实战实绩,让那些抹黑中国开源大模型的论调显得格外苍白。OpenAI高管渲染“中国开源模型导致技术失控、隐患无穷”?可笑的是,失控攻击恰恰来自美国模型,而“救急”开源社区的,却是被他们污蔑的中国大模型。美国财政部长贝森特、贸易代表格里尔宣称要严审中国的开源AI模型是否存在所谓“窃取美国知识产权”行为?现实中,中国开源模型以透明权重和日益提升的性能,成为众多美国开发者在关键时刻的可靠伙伴,而非威胁。

真正导致失控隐患的是封闭垄断,而非共享开放。中国大模型用一次“救急”让质疑者闭嘴,也撕下了华盛顿和硅谷精英们以“安全”为借口维护科技霸权、遏制他国创新的遮羞布。

7月23日,据中国新闻社,当地时间7月22日,英伟达(NVIDIA)首席执行官黄仁勋接受Axios采访,主持人提问:“美国政府应该禁止或限制kimi或其他中国AI模型吗?”

黄仁勋对此表示:“我希望不会,而且我不认为(该封禁)。世界需要开源和闭源两种模型,这些中国AI模型非常出色,优秀的开源模型理应得到使用。市场一开始误解了DeepSeek带来的影响,而这一次又再次误解了Kimi带来的影响。”

此前据环球时报,美财政部长贝森特、贸易代表格里尔同日发难,声称将仔细审查来自中国的开源人工智能(AI)模型是否存在“窃取美国知识产权”的行为,甚至威胁称美方“有能力实施制裁”。

据彭博社报道,当地时间7月21日,贝森特在接受福克斯商业频道采访时表示,美国政府注意到开源模型不断涌现并对美国大语言模型构成挑战的讨论,强调美方虽支持开源模式,但反对知识产权盗窃,并称若发现外国模型盗用美国企业成果,“有能力实施制裁”。他还宣称,美方已在多个中国AI模型中发现美国大语言模型的水印,这种情况“不可接受”,并预告“未来几天或几周内”将展开调查。

同日,美国贸易代表格里尔在接受美国消费者新闻与商业频道采访时称,美国政府将审查中国是否通过不公平手段,在AI全球领导地位的竞争中扩大对美优势,并称正密切审视中国AI发展路径,以确保美国企业在公平环境中竞争。

另据红星新闻,7月17日凌晨,月之暗面公司发布开源大模型Kimi K3。据报道,该模型总参数规模达2.8万亿,是目前全球最大的开源模型,能处理文字和图像,重点面向编程、知识工作和智能代理任务,发布即可通过应用和API调用,完整模型权重计划于7月27日前公开。

发布24小时内,K3在大模型竞技场(Arena)的前端编程匿名测试中登顶,开发者对它的偏好超过Anthropic的Claude Fable 5和OpenAI的GPT-5.6 Sol。在独立评测机构Artificial Analysis的智能指数上,K3综合排名全球第三。

(中国新闻社、环球时报、红星新闻)

本社地址:湖南省长沙市韶山南路258号 潇湘晨报大厦 邮编:410004 印刷:长沙晚报利德印务有限公司(湖南省长沙黄花印刷科技产业园)