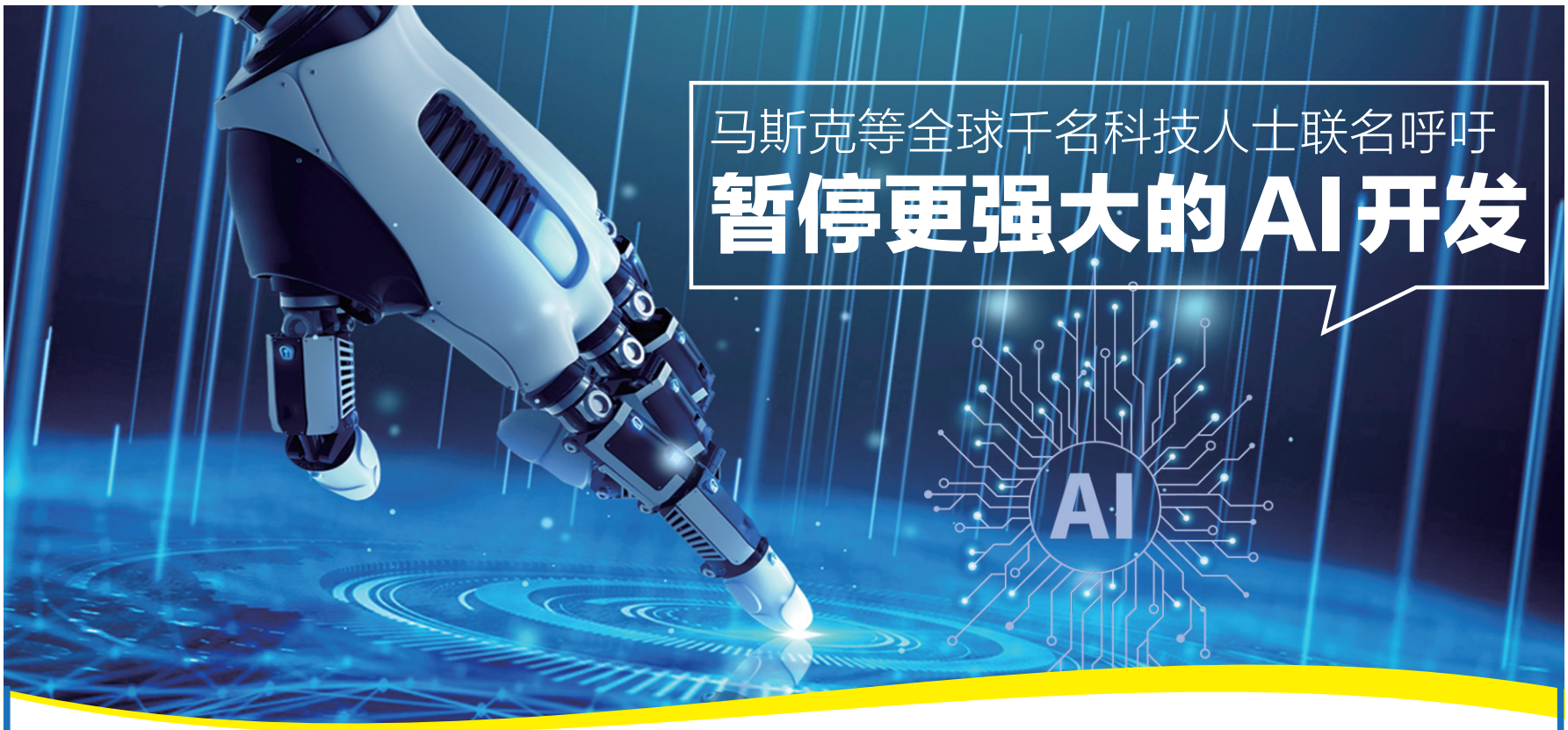

马斯克等全球千名科技人士联名呼吁

暂停更强大的AI开发

最近，有关ChatGPT和人工智能，几乎每天都有新闻。随着更“聪明”的GPT-4发布，人们对技术的恐慌和对科技伦理的担忧，得到了更多关注。

最新的消息是，当地时间3月29日，美国非营利组织未来生命研究所发布了一封名为“暂停巨型AI实验”的公开信。上千名人工智能专家和行业高管在信中呼吁，所有人工智能实验室应当暂停对更强大的人工智能系统的开发和训练。至少，暂停半年。他们建议，如果不能迅速实施这种暂停，“政府应介入并实行暂停”。

人工智能对社会和人类的潜在风险，已经成为不少科技人士的共识，其中包括“人工智能教父”杰弗里·辛顿、特斯拉和推特CEO埃隆·马斯克、图灵奖得主约书亚·本希奥。为此，他们在这封公开信上签下了自己的名字。

“只有当我们能够确信，AI体系的有利因素都是积极的，且风险都可控，一个强大的AI体系才能成熟。”公开信中这样写道。

实际上，这不是未来生命研究所第一次公开呼吁对人工智能发展的警惕。2014年，这个组织在美国成立，成立的初衷一方面是促进对“未来乐观图景”的研究，一方面则是“降低人类面临的现存风险”。而后者一直是其关注的重点。

2015年，物理学家斯蒂芬·霍金和埃隆·马斯克等数位科学家、企业家，与人工智能领域有关的投资者，联名发出了一封公开信，警告人们必须更多地注意人工智能的安全性及其社会效益。

那时，AI还没有像今天这样，呈现出令人不安的“智能”。但从那时候开始，马斯克就已经表示，他坚信不受控制的人工智能“可能比核武器更危险”。

8年后，在愈来愈动荡的经济形势中，新一封公开信的签署者们提出疑问：“我们是否应该让机器用宣传和谎言充斥我们的信息渠道？我们是否应该自动化所有工作，包括令人满意的工作？我们是否应该发展最终可能比我们更多、更聪明，淘汰并取代我们的非人类思维？我们应该冒险失去对我们文明的控制吗？”

欧洲刑警组织3月27日也发出警告，ChatGPT等人工智能聊天机器人很可能被犯罪分子滥用：“大型语言模型检测和重现语言模式的能力，不仅有助于网络钓鱼和在线欺诈，还可以用来冒充特定个人或群体的讲话风格。”他们的创新实验室已经组织了多次研讨会，探讨犯罪分子可能会怎么做，列出了潜在的有害使用方式。

这些签署者希望，奔向危险的脚步能够“暂停”，先共同开发出一套用于高级人工智能设计和开发的共享安全协议，并由独立的外部专家进行严格审计和监督。AI开发人员也需要与政策制定者合作，开发强大的AI治理系统。

据《中国青年报》

深度分析

AI生成的合成图像泛滥且真假难辨

政策监管势在必行

近日，美国前总统特朗普被全副武装的纽约防暴警察按倒在地的图片在推特等社交媒体平台泛滥，但这些看似细节丰富的图片却与事实毫不相干，这些图片出自人工智能(AI)驱动的图像生成技术。

专家们警告说，这些图片昭示出一个新现实：在重大新闻事件发生后，虚假图片和视频有可能充斥社交媒体，进一步混淆事实，因此亟须部署相关技术并制订相关政策，对类似技术进行监管。

据《科技日报》

合成图像真假难辨

据美国《财富》杂志网站报道，这些“特朗普被捕”的图片由总部位于荷兰的开源调查媒体“响铃猫”网站的创办人艾略特·希金斯生成。

当希金斯看到有关特朗普可能被捕的消息时，决定将其可视化。为此，他使用最新AI绘画工具Midjourney制作了关于特朗普被捕的图片。他表示，最新工具比之前的版本复杂得多，极大地改进了图像的视觉效果。随后他在推特上分享了合成结果：前总统被警察包围的图片，其中徽章做了模糊处理。据美国《华盛顿邮报》网站报道，仅两天，希金斯发布的这条帖子被浏览了近500万次。

人工智能专家表示，虽然处理并生成虚假图片的技术并不新鲜，但该领域技术的进展速度以及人们对技术的滥用值得关注。数字内容分析公司Truepic的穆尼尔·易卜拉欣指出，“合成内容正在快速发展，真实和虚假内容之间的差距变得越来越难分辨”。

《财富》杂志指出，如今各种AI图像生成工具变得触手可及，它们可在用户发出简单指令后，迅速生成海量栩栩如生的图片。例如，Midjourney这款文本到图像模型现在可生成模仿新闻机构照片风格的图像，如此一来，这些AI生成的图片有望在混乱的新闻环境中“浑水摸鱼”，混淆视听。

专业人士强调称，大量制作虚假但看似可信的图像的能力已有了巨大进步，而且很容易用于欺骗目的。

技术打击“深度伪造”

希金斯认为，随着合成图像越来越难辨真假，对抗错误视觉信息的最佳方式是提高公众的意识并加强这方面的教育，社交媒体公司可把重点放在开发能够辨别AI生成图像的新技术上，并将这种技术融入自己的平台。

比如，推特出台了相关政策，禁止用户分享可能造成伤害的欺骗性和操纵性媒体内容，例如可能导致暴力、广泛内乱或威胁个人隐私的推文。今年2月，推特推出了“社区笔记”功能，允许用户在推文下方添加注释，并且以长内容的形式进行解释。有观点认为，“社区笔记”或将能够帮助推特成为更可信、让更多人积极发言的平台，并帮助减少假消息、误导性内容的比例，以此吸引更多广告投放。

政策监管势在必行

“特朗普被捕”虚假图片在互联网泛滥还提供了一个案例研究，表明目前缺乏企业标准或政府法规来解决使用AI制造和传播谎言的问题。

纽约研究人员米歇尔说，他担心这个世界还没有准备好迎接即将到来的铺天盖地的虚假信息。

专家们一致认为，特朗普的名气使虚假图片很容易被发现，但识别出普通人相关的虚假图片可能困难重重，而且生成虚假图片的技术一直在进步。从政策角度来看，亟待以立法形式对深度合成技术的应用进行规制。

新闻延伸

人工智能(AI)

人工智能(Artificial Intelligence)，英文缩写为AI。它是研究、开发用于模拟、延伸和扩展人的智能的理论、方法、技术及应用系统的一门新的技术科学。

人工智能是计算机科学的一个分支，它企图了解智能的实质，并生产出一种新的能以人类智能相似的方式做出反应的智能机器，该领域的研究包括机器人、语言识别、图像识别、自然语言处理和专家系统等。人工智能从诞生以来，理论和技术日益成熟，应用领域也不断扩大，可以设想，未来人工智能带来的科技产品，将会是人类智慧的“容器”。人工智能可以对人的意识、思维的信息过程的模拟。人工智能不是人的智能，但能像人那样思考、也可能超过人的智能。