捏造公众人物讲话 参与军事行动应用

人工智能风险骤增如何规制？

文 / 艾舟

过去的一周，人工智能技术正以难以想象的方式嵌入到人们的认知与生活里：从频频模仿公众人物讲话，到实际参与血腥残酷的军事冲突，人工智能让越来越多的人感受到何谓“一念成佛，亦可一念成魔”。

随着首届全球人工智能安全峰会举行，如何有效规避这一新兴技术的风险，避免打开智能革命时代的“潘多拉魔盒”，已迫在眉睫。

■ 首届全球人工智能安全峰会在英国召开 图GJ

■ 人工智能生成视频已不是新鲜事 图GJ

模仿政要以假乱真

一个“日本首相岸田文雄”的视频上周在全球社交媒体疯传。

视频中的“岸田”，身着西装，语气郑重，屏幕上还配以日本电视台的台标和“突发新闻”字样。然而，从他嘴里讲出的，却是不堪入耳的“黄段子”。

这当然是一则生成式人工智能制作的假视频。但惊人的是，视频在社交平台上发布后，短短24小时播放量已超过230万次，关注度与传播率可谓惊人。

更惊人的是，据制作该视频的男子表示，生成这则视频的原料其实就是日本电视台新闻节目的实际影像，使用工具则是能够配合说话声音进行嘴型加工的生成式人工智能软件。据悉，这一软件学习岸田的声音用了约2小时，制作视频本身则只用了约1小时。

所谓生成式人工智能技术，是指具有文本、图片、音频、视频等内容生成能力的模型及相关技术。换句话说，它可以创造原本不存在的对话、故事、图像、视频和音乐，模糊真实与虚假的边界，令人难以判断。

如果说生成式人工智能所创作的作品是否享有版权还主要是法律问题，对公众人物的言论近乎以假乱真的杜撰则不仅涉及伦理和法律问题，更有可能引发难以预料的后果。尽管制作“岸田”视频的男子称自己只想搞笑，且为此向日本电视台表示歉意，但细想难免令人后怕——一国政治首脑若可以被人操纵面向公众发言，这次是“开黄腔”，那下次如果是对外宣战呢？

其实，早在“岸田”讲段子前，不少政治人物已成为生成式人工智能创作的原料。这一技术还可能在选举的关键节点生成诋毁竞选对手的假视频，从而以假消息左右选举。也难怪有分析人士指出，生成式人工智能的发展，有可能对目前的选举政治构成前所未有的致命挑战。

政坛之外，生成式人工智能对娱乐、传媒、艺术等领域公众人物的“创作”也不少见。在各种短视频平台，从以搞怪喜剧闻名的“憨豆先生”，到因饰演《哈利·波特》系列电影里的赫敏而成名的艾玛·沃森，不少外国名人纷纷能“讲出”一口毫无翻译腔、又完美符合本人音色的普通话。

如果说这种“全世界都在说中国话”的“创作”还只是逗人一乐，那么生成式人工智能依据人物头像等资料合成制作的色情制品，则是对相关人物赤裸裸的侵犯与伤害，会对当事人造成困扰，乃至巨大的名誉和精神创伤。

更有甚者，生成式人工智能还被用于电信诈骗。作为一家科技公司法人代表的郭先生突然接到一名“好友”的微信视频，称他在外地竞标，需要430万元保证金，想借用郭先生公司的账户走账。基于对好友的信任，加上已经视频聊天“确认”了身份，郭先生在10分钟内把430万元巨款如数转到了对方的银行账户上。事后郭先生才得知，这其实是骗子通过人工智能换脸和拟声技术实施的骗局。

美国电视台也曾做过一期节目，现场展示黑客通过克隆声音来窃取信息的过程。黑客仅用了5分钟就通过人工智能软件克隆了当事人的声音，并且替换了手机的来电显示功能，成功获取了相当多的私人信息。

军事应用引发担忧

在亚欧大陆另一端，以色列与巴勒斯坦伊斯兰抵抗运动（哈马斯）在加沙地带血战的背后，也惊现人工智能的身影。

以色列国防军一名高级情报官员透露，自10月7日本轮巴以冲突爆发以来，以色列国防军借助人工智能打击了加沙地带1.4万多个目标，“高度先进的人工智能主导的目标库”帮助以军士兵在一天内锁定并摧毁了150个地道目标。

哈马斯自2007年完全夺取加沙地带的控制权后，便开始大规模营建地道设施。1300多条地道可以有效规避以色列的侦察与打击，从而成为加沙地带地下物资流转与储备粮弹、制定战略战术的重要依托。有消息称，这些地道遍及加沙各地，地下深约70米，总长500到800公里，可以直达以色列修建的隔离墙。

即便以军高技术装备众多，但在深邃复杂的地道里也会失灵。在这些大多只有一肩宽的地道里，以军被哈马斯伏击的风险很大，因此以军通常禁止普通地面部队进入地道。在地道内外，哈马斯往往还会设置很多诡雷，令以军头疼不已。

在2014年那一轮持续50余天的巴以冲突中，以色列同样对加沙地带实施了大规模空袭，但破坏的隧道仅有32条，且付出至少60余名士兵阵亡的代价。而据以军高官透露，本轮巴以冲突地面战开始后不到一周，以军就成功摧毁了哈马斯十分之一以上的地道，自身伤亡也大幅降低。

实际上，以军之所以进攻效率明显提高，远非借助由人工智能主导的目标库这么简单。此次进攻地道之初，以色列就投入了大量机器人、无人机等智能设备。例如，在进入地道之前，以军士兵往往先投入手抛式侦察机器人和小型卡车机器人。这些机器人借助搭载的人工智能软件，不仅能掌握地道的位置和大小、描绘地道的线路，还往往能提前引爆诡雷，并发现和引导攻击地道内的哈马斯目标。

可以说，在本轮巴以冲突中，人工智能对战事的介入程度又达到新的水平。它发挥的作用已不只是辅助人脑进行图像识别，还直接参与到复杂地形条件下的地图绘制与侦察。从战争实践的角度来说，以军的相关经验必然引起各方重视。

然而，可怕之处也恰恰在于此。人工智能已经一步步从战场幕后走向前线，而鉴于西方在无人设备上侦察和打击一体技术的发展，人工智能芯片算力的提升，尤其是地缘政治紧张态势加剧的高技术竞争，在下一次冲突中，是否有可能看到类似电影《终结者》中自主杀人机器人的出现？

以色列宣布将人工智能融入致命性行动，距离本轮巴以冲突不过短短几个月时间。科技威力的发展如此迅速，以致让人担忧能否有效规制它。人工智能的“潘多拉魔盒”能否避免被彻底打开，实质上取决于国际社会的共识与行动。

国际监管依然缺位

首届全球人工智能安全峰会本月在英国布莱奇利园召开，来自全球28个国家的政府代表、7个国际组织，以及超过80个学术研究机构、企业和公民组织与会。

尽管与会国家数量似乎有限，但在人工智能的“玩家群”，可以说是强手尽出。除来自中国、美国、英国、欧盟、印度的官方代表外，还有英国图灵研究所、中国科学院、经济合作与发展组织等众多智库机构，以及英国“深层思维”公司、美国谷歌公司和微软公司、中国腾讯公司和阿里巴巴公司等知名企业参与。本次峰会的重中之重，就是关注人工智能技术可能带来的风险，探寻全球人工智能治理方案。

对于这一幕，百岁高龄的美国前国务卿基辛格可能会感到欣慰。这位曾推动东西方之间架起沟通桥梁的外交家，近年来一直关切人工智能对全球未来的冲击，并大声疾呼大国之间应合作管控新兴技术的科技风险。

然而，就在此次峰会开幕之前，一些西方政客却对是否应邀请中国与会提出异议。对此，英国首相苏纳克在宣布邀请中国的决定时表示，就人工智能相关科技实力而言，美国居首，中国紧追其后，若不让全球人工智能强国参与研讨过程，就不会有值得被严肃看待的全球人工智能战略。

事实上，一些西方政客动辄以意识形态为由竖起科技藩篱的做法，恰恰是阻碍全球人工智能治理的最大阻碍之一。尽管峰会发布了《布莱奇利宣言》，但距离形成具有国际法约束力、被各方均认可的人工智能治理规章尚有不小距离。

值得关注的是，10月“一带一路”国际合作高峰论坛期间，中国发布了《全球人工智能治理倡议》。其中指出，各方应鼓励全球共同推动人工智能健康发展，共享人工智能知识成果，开源人工智能技术；反对以意识形态划线或构建排他性集团，恶意阻挠他国人工智能发展。中国还提出，积极支持在联合国框架下讨论成立国际人工智能治理机构，协调国际人工智能发展、安全与治理重大问题。

纵观历史，科技的发展从来都是双刃剑，但剑刃指向哪里可以由人来决定。在人工智能迭代加快、应用愈发广泛的当下，各国亟需共同行动，防止人工智能“野蛮生长”。