瑞莱智慧：用AI让AI更安全

从ChatGPT火爆全网，到“AI换脸”新骗局全国频发，人工智能作为战略性新兴技术受到了市场广泛关注。如今，在“AI换脸”诈骗多发的背景下，如何破局AI技术滥用、如何应对AI技术安全风险无疑也引发业界深思。5月24日，瑞莱智慧RealAI合伙人、高级副总裁朱萌接受了北京商报记者专访，从目前的“AI换脸”骗局如何防范、鉴别，到其背后的深度合成技术，再到推动建立安全可控的AI治理体系，朱萌一一详谈了她的看法。

信息造假难度降低

“张嘴、摇头，摄像头捕捉到人的动作后，屏幕上原本静态的人像可以动起来，动作幅度和真人一致，还能眨眼和露齿微笑，仿真度颇高，几分钟内就能生成一段视频，脸还能被任意替换。”这就是“AI换脸”技术，成为不少黑产分子新型诈骗的工具。

近期，一家公司老板被“AI换脸”10分钟骗走430万元引发市场热议，“AI换脸”骗局不断。其中，包头市公安局电信网络犯罪侦查局近日也发布一起使用智能AI技术进行电信网络诈骗案件，不法分子利用“AI换声”“AI换脸”技术，伪装成特定人物，实时与他人进行视频通话。更换后的面部表情自然，以假乱真，能够冒充他人身份联系被害人，博取被害人信任后实施诈骗。

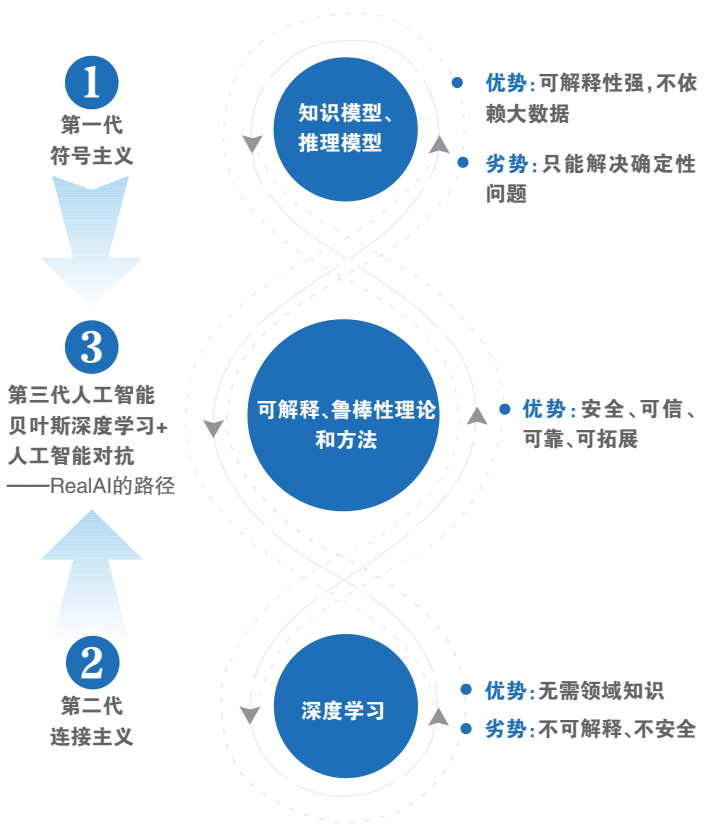
据了解，“AI换脸”背后采用的是深度合成技术，利用深度学习、虚拟现实等生成合成类算法，制作图像、音频、视频等信息，目前在社交、影视、广告、医疗等诸多领域不断深化应用，有较大的技术价值和商用潜力，不过也存在着一定的安全隐患，“双刃剑”效应明显。

朱萌在接受北京商报记者专访时表示，随着研究的深入，深度合成技术得到快速发展，早期的换脸视频，由于技术不成熟、不完善，尚存在“微表情不自然”“面部边缘有锯齿”等明显换脸痕迹，可作为参照辨别真假。但近年来在社交媒体上广泛传播的换脸视频，动作的逼真度、自然度，以及视频整体的清晰度、流畅度都得到大幅提升，足以达到以假乱真水平，传统基于生物特征的鉴别方式越来越难以发挥作用，真假难辨的背后更是危害性的不断增强。

朱萌进而指出，深度合成技术大大降低了信息造假的难度，只需要拿到一张照片，就能生成非常逼真的伪造视频，用于捏造虚假信息、伪造不雅视频、恶搞特定人物等。

5月24日，中国互联网协会也发文称，面对利用AI技术的新型骗局，广大公众需提高警惕，加强防范。

第三代人工智能——瑞莱智慧技术路径



一道安全防火墙

针对“AI换脸”等技术被用于诈骗、诽谤等问题，朱萌对北京商报记者表示，RealAI已开发配套的治理工具，即深度伪造内容检测平台DeepReal，为防范大规模的视频造假提供技术支撑。

目前，学术界和产业界均已对深度合成鉴别检测技术的研发进行了大量投入。在国内，清华大学、中科大等高校均在深伪检测方面取得显著成果。瑞莱智慧RealAI、腾讯优图实验室等企业机构也已构建人脸合成检测平台并发布针对性的检测产品，支持对多种换脸方法进行检测。

据朱萌介绍，RealAI推出的深度伪造内容检测平台DeepReal，通过辨识伪造内容和真实内容的表征差异性、挖掘不同生成途径的深度伪造内容一致性特征，能够快速、精准地对图像、视频、音频内容进行真伪鉴别，有效打击财产诈骗、虚假宣传、证据造假等违法行为。

据了解，DeepReal平台基于千万级训练数据，能够实现毫秒级检测速率，每帧画面检测时间仅需30毫秒，同时在主流数据集检测准确率达99%以上，在实际产业中检测准确率也达到业界顶尖水平。

深度合成技术的加速应用已成为现实趋势，据朱萌介绍，RealAI未来还将进一步扩展深度合成溯源、深度合成鉴定等方面的能力，全面提升技术应对能力。

目前，瑞莱智慧的产品解决方案主要在政务、金融、能源、制造、互联网等场景落地，致力于以安全、可信、可靠、可扩展的第三代人工智能技术为合作伙伴提供人脸识别系统安全、自动驾驶系统安全、深度合成和伪造检测、隐私保护计算、AI攻防靶场等全套产品和解决方案。

北京商报记者了解到，除了瑞莱智慧RealAI之外，部分涉足人脸识别的上市AI公司也已研发了视频图像鉴真相关产品。

让AI安全可控

伴随着“AI换脸”骗局在全国范围内的爆发，AI技术的安全问题也受到了广泛重视。今年以来，有关推动建立安全可控的AI治理体系也被不断提及。

行业调查上，德勤今年发布的《2023技术趋势》报告显示，有七成左右的受访企业主管表示将看好AI技术业务变革；47%的业务主管更关心AI透明度问题；而41%的技术专家担忧AI会引发道德风险。

朱萌表示，包括人脸识别、深度合成在内的人工智能技术的普遍应用，可能引发诸多治理挑战。从整个大市场来看，AI治理应从法规与倡议约束、研究计划与竞赛引导、技术手段发展的角度全面展开。其中技术手段发展方面，应对AIGC生成内容的不断演化，对治理技术也提出了越来越高的挑战。从数据制作、传播到事后检测追责的全链条看，需要技术侧不断优化合成检测、检测结果可解释等方面的技术能力。国内外相关机构也均推出和优化相应的技术工具。

据了解，为推动人工智能伦理治理体系的建立，RealAI在2022年6月正式成立AI治理研究院，搭建了由“研究体系”“实践体系”和“监督体系”组成的治理框架，持续开展人工智能伦理规则、治理实践与监督体系的研究，着力于伦理规则与立法贡献、治理技术落地与探索伦理安全规范。

此外，据朱萌介绍，推动建立负责任的AI技术治理体系，RealAI主要从“算法可靠、数据安全、应用可控”三个方面着手，提升算法模型的可靠性，解决模型训练与使用过程中的隐私保护与技术滥用问题等，还搭建了安全可靠可控的新一代人工智能基础设施，且目前已在政务、金融、工业互联网等高价值场景中发挥了重大作用，成为AI纵深赋能的坚实基座。

北京商报记者 马换换