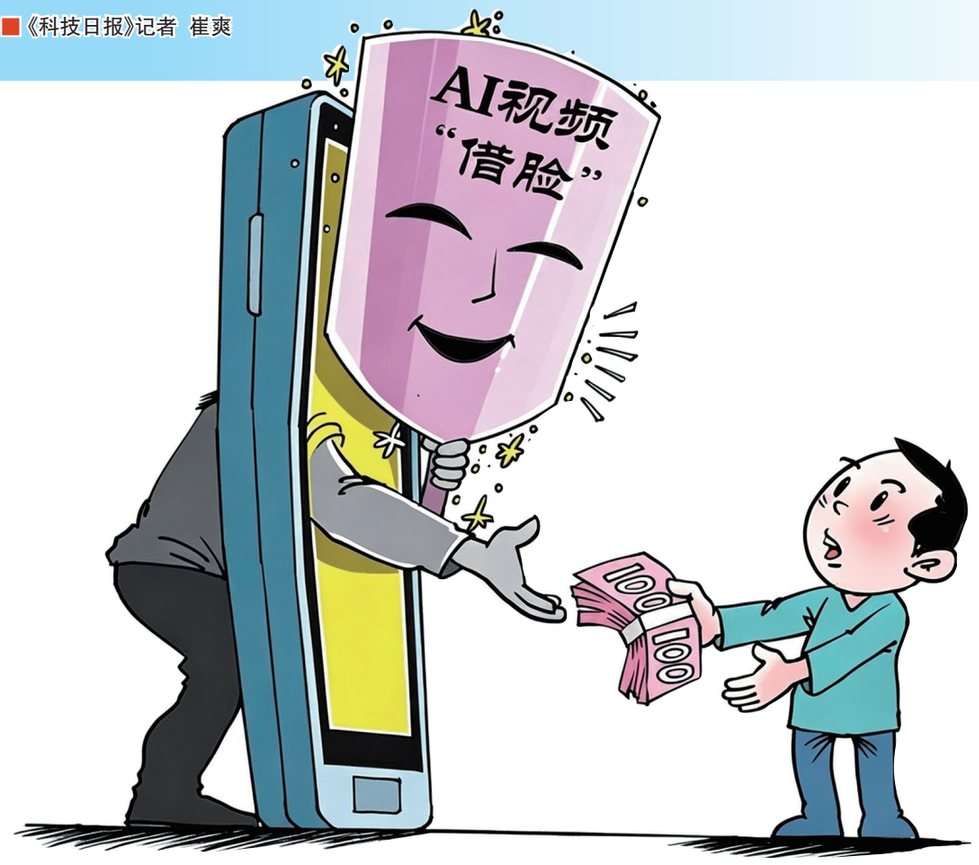
AI生成视频已能骗过人脸识别！安全防护该如何升级？

■《科技日报》记者 崔爽

近日，杭州市中级人民法院公布一起案例：两名不法分子利用网络收集他人照片，通过AI生成点头、摇头、眨眼等动态视频，再通过篡改相机数据，成功突破购物平台的人脸核验，盗取账户内个人信息。

AI为何能骗过系统？在AI换脸突破主流核验的情况下，人脸识别是否安全？这种风险又该如何来应对？科技日报记者就此采访了有关专家。



图/余宁山

第一问
“AI脸”与真人脸为何难以区分？

“传统人脸核验主要依靠活体检测，通过点头、摇头、眨眼等动态动作，区分真实用户与静态照片、视频。”天翼安全科技有限公司高级专家危嘉祺对记者解释，“但随着生成式人工智能技术快速发展，AI已经能够根据单张照片生成高度逼真的动态人脸视频，甚至模拟符合核验要求的动作。部分攻击还可结合虚拟摄像头、视频流注入等技术，绕过真实摄像头采集环节，使系统难以有效区分真实人脸与合成内容。”

但这并不意味着人脸识别技术已经失效，说明依赖单一动作验证的传统技术路线面临新的安全挑战。“人脸识别仍具有重要应用价值，但不能将其等同于绝对可靠的身份认证手段。”他说。

第二问
人脸核验相关风险由谁承担？

“AI发展带来的新型安全风险，需要坚持依法治理、权责一致、多方协同的原则，构建政府监管、企业履责、行业自律、社会监督、公众参与的治理格局。”危嘉祺表示。

具体来看，首先，对于利用AI伪造他人人脸信息、实施身份冒用、非法获取个人信息等违法犯罪行为，应依法追究相关主体的法律责任，从源头遏制技术滥用。

其次，应用平台和技术服务提供方应按照各自职责，依法落实个人信息保护和网络安全相关要求，持续完善身份核验机制，加强对深度伪造、视频流注入等新型攻击的风险识别与安全防护，切实保障用户合法权益。

最后，监管部门应进一步完善相关标准规范和安全评估机制，推动行业建立风险信息共享、漏洞通报和协同处置机制。公众也应增强个人生物识别信息保护意识，避免随意公开或向不可信渠道提供敏感信息。

“需要强调的是，不能因为新技术存在被滥用的风险，就否定技术创新和应用；也不能将技术安全风险简单转嫁给普通用户。”危嘉祺说，应坚持发展与安全并重，通过技术创新、制度完善和多方协同，不断提升AI应用的安全性与可信度。

第三问
AI安全治理面临哪些新要求？

在危嘉祺看来，AI生成视频突破人脸核验，反映出AI安全治理正在从传统的静态风险防范，转向生成技术与检测技术持续迭代的动态攻防对抗。为此，需要从监管机制和技术能力两个层面同步提升。

危嘉祺建议，在监管层面，应进一步健全人工智能安全治理体系。针对身份认证、金融交易、个人信息查询等高风险场景，推动建立与业务风险相适应的分级分类安全要求，避免过度依赖单一生物识别手段。同时，继续完善深度合成内容标识、生成内容溯源和安全评估机制，推动建立新型攻击样本共享、风险通报和动态测评制度，提升监管标准对新技术、新风险的适应能力。

在技术层面，则应坚持以攻促防，推动安全防护从被动检测向主动对抗升级。一方面，利用先进生成式AI模型构建高逼真度攻击样本，持续开展人脸核验系统的红队测试和对抗评估，及时发现防护薄弱环节；另一方面，综合运用设备可信验证、视频流注入检测、生理活体特征分析、多因素身份认证及业务行为风控等技术，构建多层次、纵深式安全防护体系。

“特别需要关注的是，安全测评不能只针对已知攻击样本，更要检验系统面对最新生成模型、未知伪造手法和复杂攻击组合时的防护能力。”危嘉祺提醒，应推动建立覆盖攻击生成、风险发现、防护验证和持续优化的动态闭环，将真实案例及时转化为安全测评场景，不断提高人工智能安全防护的实战能力。

评论

“AI盗脸”案暴露平台核验代差

杭州中院近期公布的一起案件，戳破了很多人对“刷脸安全”的幻觉：两名不法分子用网络收集来的照片，通过AI生成点头、眨眼、摇头等动态视频，再用虚拟相机替换摄像头输入，绕过购物平台活体核验，非法登录近30个账号，盗取并售卖个人信息，还违规提供虚假实名认证。法院判令二人公开道歉，并支付公益损害赔偿金5857元。

金额不大，但信号很强。过去人们常说，密码可能泄露、验证码可能被拦截，脸总偷不走。如今，这张“最像自己的钥匙”也可能被复制。问题不只是AI伪造技术变强了，更是部分平台核验机制升级慢了。依赖“指定动作”判断真人的传统活体检测，面对AI合成视频和视频流注入，已经显得单薄。

更值得追问的是平台责任。有的平台匆匆上线人脸核验，却把低成本的单一动作校验当成主要防线；有的明知虚拟相机、视频流注入已是新型攻击手法，却不能及时升级；还有的核验能力来自第三方服务商，自身不掌握底层算法，出事后容易互相推诿。平台不能因为采购了第三方服务，就免除个人信息保护和安全防护责任。

制度层面并非没有依据。《人脸识别技术应用安全管理办法》明确，能实现同等目的时，不得把人脸识别作为唯一验证方式；处理人脸信息前应开展个人信息保护影响评估，达到一定规模还需备案。最高法今年发布的涉人工智能纠纷意见也明确，未经同意利用AI生成可识别自然人的虚拟形象并使用的，构成人格权益侵害。这意味着，刷脸不能只图方便，更应按高风险个人信息处理活动来管理。

平台和技术服务方必须主动进化。核验从“单次动作过关”升级为多信号交叉验证，在高逼真度AI攻击样本基础上开展红队测试，叠加设备可信验证、视频流注入检测、生理活体特征分析和多因素身份认证。监管部门继续严查AI伪造人脸黑灰产链条，推动攻击样本共享、漏洞通报和动态测评，让伪造技术可识别、可追溯、可追责。

对个人而言，也要更新安全认知：不随意向不可信渠道提供高清正脸照、证件照和动态视频；支付、登录等敏感场景尽量开启短信、邮箱、设备锁等二次验证；发现异常登录及时冻结账号并留存证据。

刷脸、眨眼，本意是为了让人们更安心地享受数字生活。但技术越便利，越不能把安全寄托在“看起来像本人”的单一判断上。面对AI深度伪造的加速演进，平台要升级、监管要跟进、个人要警惕，共同把数字生活的安全底线扎得更牢。

（来源：人民网）

案例

AI换脸盗刷案敲响个人信息安全警钟

数字时代，人脸识别技术带来了生活便利，但也面临风险，“人脸信息”一旦泄露，可能比密码泄露更可怕。今年3月，湖南省长沙市雨花区人民法院披露一起AI“换脸”盗刷银行卡刑事案件。

2025年2月，面临大学毕业的吴某没有选择正规就业，反而盯上了网络黑灰产业，并很快发现了“商机”——用户“张哥”正在兜售一项违规技术，能将静态人脸照片转化为可通过某支付软件人脸识别验证的AI动态视频。

吴某随即购买了使用该技术的账户，随后主动在加密聊天平台上发布广告，吸引诈骗团伙找他“下单”。据悉，诈骗团伙往往冒充“抖音客服”或“游戏主播”，以“关闭自动扣费”或“赠送游戏装备”为诱饵，诱导受害人下载远程控制软件，获取其人脸、身份证、银行卡号、手机号等信息。

利欲熏心的吴某联络“上线”并接收被害人的手机号码、照片等个人信息，制作被害人的人脸AI视频，以此登录被害人的云闪付账号并修改密码，随后交给“下线”人员盗刷被害人云闪付所绑定的银行账户资金。其间，被害人张某、王某、李某的银行账户均以上述方式被盗刷人民币总计5万余元。2025年8月，吴某经民警电话联系主动到案，其家属退赔了被害人的经济损失。

经审理，雨花区法院综合考虑吴某具有自首情节，认罪悔罪表现及对被害人的经济赔偿等情况，最终判处吴某有期徒刑1年10个月，并处罚金3000元。

法官表示，本案折射出技术快速演进背景下，数字安全、行业监管与个人素养等多维度的治理短板。比如，单一人脸识别的验证模式存在漏洞，多重生物识别与行为特征验证的复合防护机制，应当成为优化的方向。又如，公众的个人信息保护意识亟待强化。社交平台上随意发布的人脸照、身份证信息，可能成为AI的伪造“原料”。公众应审慎发布敏感生物信息，养成“最小必要披露”的信息保护习惯。

延伸阅读

主要风险

- **AI换脸盗刷：**这是目前最凶险的威胁。不法分子利用AI技术，将人的静态照片生成动态视频，直接突破人脸识别系统，盗刷你的账户资金。
- **强制收集与滥用：**部分商家强制要求“刷脸”作为唯一验证方式，甚至通过“无感抓拍”收集人脸数据，用于精准营销或“大数据杀熟”。
- **数据泄露与关联：**人脸数据一旦泄露，不法分子可以将之与你的身份信息关联，重构你的“关系网”和行为轨迹，甚至用于洗钱等犯罪活动。

应对方法

- **非必要不提供：**对于非核心业务（进小区、取快递），如果提供身份证或手机号也能完成验证，尽量拒绝“刷脸”。
- **警惕“免费”陷阱：**不要为了小礼品（如鸡蛋、洗衣液）轻易授权刷脸，这往往是非法收集信息的陷阱。
- **启用二次验证：**在涉及资金交易的高风险场景，务必开启“人脸+密码”或“人脸+短信”的双重验证，增加安全屏障。
- **关注新规：**2025年6月1日起施行的《人脸识别技术应用安全管理办法》规定，不得将人脸识别作为唯一验证方式，你有权拒绝非必要的“刷脸”。

（来源：长沙晚报）

责编/张会 版式/李昂 校对/李雅

