AI换脸拟声威胁金融防线 机构“深伪”对抗全面升级

本报记者 李晖 北京报道

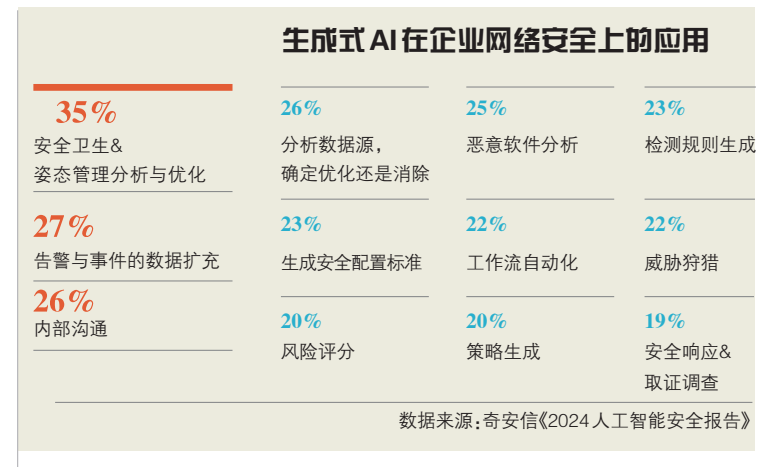
技术大爆发永远是一把双刃剑。随着生成式人工智能技术的发展和普及，技术门槛大幅降低，也为不法分子提供了机会。

2023年以来，“通过AI换脸和拟声技术实施诈骗”的新型骗局频繁登上新闻头条，引发大众关注。在机构侧，数据显示，在全球范围内，有接近一半（46%）的企业遭受过合成身份的欺诈，90%的受访企业认为这种行为已日益严重。

金融行业作为资金汇聚地一向是不法攻击的主要目标。中国信通院人工智能所安全与元宇宙部工程师邹皓在接受《中国经营报》记者采访时表示，虽然深度伪造（Deepfake，以下简称“深伪”）攻击在金融机构遭到的网络攻击中绝对占比可能不是最高，但随着AIGC发展，生成一段伪造视频的技术门槛和所需资源越来越低，有可能将在下一阶段成为威胁金融行业安全的重要因素。

“用魔法打败魔法”是对抗技术攻击的核心要义。记者近期采访了解到，当前国内金融行业在防御人脸和声音深伪方面的投入不断加码，部分金融机构开发了防深伪的检测模型以应对这一新型威胁，并加大相关领域团队建设，而一些领先金融科技公司也开始将相关技术能力产品化。

一家国有银行人脸业务项目负责人向记者表示，检测技术有一定的滞后性，新的算法需要样本积累和过滤，机构自身能做的就是争取在犯罪前捕捉行为动机，无限贴近作案时点。



深伪渐成金融业重要威胁因素

在已开展的移动端评测中，人脸识别产品首次送检被攻破的概率高达71%，二次送检被攻破的概率也有25%。

人脸识别技术以其高效的身份验证特性，在金融行业中广泛应用，这也导致相关领域面临的基于AI的人脸攻击案例开始抬头。

公开信息显示，2021年，一家大型银行受到来自IP地址为中国台湾的黑客攻击，该攻击7次通过了该行的人脸识别、6次通过了活体检测，最终导致多位储户损失合计数百万元。

记者从一家第三方评测机构获得的数据显示，在已开展的移动端评测中，人脸识别产品首次送检被攻破的概率高达71%，二次送检被攻破的概率也有25%。

AI换脸主要是使用深度合成技术。邹皓告诉记者，其技术逻辑是使用大量的人脸数据，通过深度学习算法和神经网络，训练模型识别理解人脸的关键特征。在换脸的过程中，使用训练好的模型提取原始人脸照片或视频的特征，然后将目标人脸与之相匹配，实现这些特征的转移。

“前几年，实现AI换脸存在一定的门槛，因为通常需要高质量的数据集对模型进行训练并保证足够的算力资源支持。但近年来相对成熟的深度合成开源工具，让AI换脸的成本逐步降低，速度和拟真度不断提升，目前几分钟到几小时就能制作一段肉眼难以辨别的换脸视频。”邹皓直言。

金融领域相比于个人用户，网络攻击更为复杂和高风险。相比传统的呈现式欺诈攻击（照片、电子屏、面具等为主），近年来更加复杂的注入式攻击（底层摄像头驱动窜改、传输层抓包、系统函数劫持等）甚至对抗样本攻击均在迅速增加。全球技术研究与咨询机构Garther研究数据显示，虽然呈现式攻击目前仍是主流，但注入式攻击在2023年增加了200%。

奇富科技首席算法科学家费浩峻向记者表示，在辨识度上，深伪技术能够生成非常逼真的人脸图像和视频，使得伪造内容难以被肉眼识别，增加了攻击的隐蔽性；而利用对抗样本技术，针对特定的检测模型生成能够绕过检测的伪造内容。此外，模式上的升级，除了视频和图像，深度伪造技术还可以结合声音、文本等其他模态，实现更全面的欺骗；通常还会配合一定的隐私数据，以一定脚本去构建攻击方案，更难防御。

相比国内较为完善的证件管理环境和AI基础设施，一些新兴市场面临的威胁更加触目惊心。

根据身份验证提供商Sumsub最新年度报告，2022—2023年，全球多国与深伪相关的身份欺诈案件激增，例如菲律宾的诈骗企图数量同比增长4500%。

“在东南亚地区，民众对个人信息保护的意识相对薄弱，证件管理体系尚待完善。由于肤色和五官特征与国内人群存在差异，在构建人脸识别算法的数据集时限制了样本的多样性，进而影响了人脸识别技术的准确率，”信也科技算法科学家吕强告诉记者。

他向记者表示：在过去的几年中，传统的人脸识别算法主要依赖于端到端的合成技术。而当前，大模型技术的兴起使得底层架构趋向于对抗式生成网络。这种架构在训练过程中不断地识别并针对问题进行迭代优化，显著提升了合成图像的真实性。因此，我们需要在业务扩展过程中不断进行技术的调试和升级。

技术对抗加速升级

近年来多家银行、金融科技公司都在加大应对深伪对抗的资金和资源投入。

应对深伪攻击，与杀毒和造毒一样，是长期攻防对抗的过程。

“深伪攻和防是一个相对且不断精进的过程，你在进步，深伪也在进步，我们要做的就是跑在它的前面。”蚂蚁集团旗下可信身份平台ZOLOZ产品总监陶冶向记者表示。

今年4中旬，ZOLOZ正式上线了深伪综合防控产品Deeper，实现在用户刷脸场景中有效拦截“AI换脸”风险。据陶冶透露：蚂蚁集团天玑实验室会通过GAN模型生成超30万测试样本，交给ZOLOZ Deeper进行判别训练，每个月还会对其进行超过2万次的攻防测评，模拟上百种伪造攻击情况。

有矛一定有盾，矛的特点变化也需要盾去与时俱进改变既有应对思路。

“我们正在构建一些大规模的模型来对抗深伪技术，通过精细化地捕捉到虚假语音中逻辑不一致的微妙特征。例如，合成的人脸可能存在两只眼睛的对称性问题，或者眼神的角度不一致；而在语音方面，声音的颤抖可能表明紧张情绪，或者是语音合成的产物。类似于测谎仪通过捕捉这些细节变化来测谎，我们的AI大模型也通过相似的机制来识别虚假语音。”吕强向记者表示。

近年来多家银行、金融科技公司都在加大应对深度伪造对抗的资金和资源投入。据费浩峻透露，公司持续加强对深伪技术背后算法的技术研发以及人才投入，在图像/视频，音频方向上都成立了专业的研究团队，以更好理解伪造内容的生成机制，并研发更有效的检测方法。

“目前我们最大的困扰是大量的不法中介，通过鼓动欺骗普通消费者的方法，来帮助消费者骗贷，从中获取暴利，对消费者和机构都带来了很大的伤害。”费浩峻直言。

据其透露，为应对这一挑战，团队已经开始通过多重的AI技术去识别这种行为，在对抗过程中模拟攻击手段来训练检测系统不断升级自己的识别能力，并结合业务流程和业务行为，对用户风险行为进行预估，通过升级环境监测、语音情绪连贯性、声纹、对话异常分析等多种手段，对案件进行打击。

“只提供身份安全能力是不够的，在具体的业务中需要各种终端能力的结合，比如事前用身份安全，事中通过交易过程反欺诈，事后通过反监督机制整体做拦截。”陶冶表示。

从应对效果看，ZOLOZ在印尼和菲律宾的一些客户发现深伪类攻击后切换到Deeper，目前攻击已经是清零状态。

如何“比黑客早半步”

技术发展过程当中攻防需要不断演练，漏洞悬赏奖金和赛事则是业界“化被动为主动”的安全策略。

除了机构自身的研发建设，如何通过各界合作，寻找全行业对抗“技术作恶”的合力，实现“比黑客早半步”也至关重要。

记者了解到，针对人脸识别应用的安全、合规的问题，中国信通院在2021年4月发起了“可信人脸应用守护计划”。

据邹皓透露，通过打造自动化机械臂测试环境、主动配合式机器人、自动化算法测试平台等自动化测试工具，其构建了“人脸识别安全评测实验室”，可以复现各类攻击行为。已经帮助30余家金融机构、技术企业发现人脸识别系统、声纹识别系统的安全风险。

技术发展过程当中攻防需要不断演练，漏洞悬赏奖金和赛事则是业界“化被动为主动”的安全策略。

公开信息显示，微软曾在2019年送出史上最高一笔漏洞挖掘奖励，总额高达20万美元，称发现这一漏洞为数十亿用户提供了保护。而今年亦有黑客通过发现特斯拉系统漏洞，赢得20万美元奖金和一辆Model 3。

记者了解到，今年4月ZOLOZ联合蚂蚁安全响应中心（AntSRC）设立了超百万的奖金池，支持安全极客来挖掘ZOLOZ Deeper的漏洞，通过“蚂蚁集团安全响应中心”官网提交漏洞情报。

“我们希望通过这种方式把技术滥用转化成正面力量，通过鼓励更多白帽子黑客发现问题漏洞，在不断攻防演练过程当中提升我们的技术能力。”陶冶直言。

同时，一些机构也开始在一些亟待提升的技术维度布局。比如信也科技将今年的“信也科技杯”全球人工智能算法大赛的主题定为“语音深度鉴伪”。

吕强告诉记者：在一些关键价值场景中，AI生成的语音欺诈行为日益增多。尽管如此，语音鉴伪技术的发展却相对滞后。“我们期望在比赛中看到参赛者提出更多能够识别新型假语音、特别是大模型生成的假语音的方法，并希望这些成果能够应用到实际生产中。比赛结束后我们会开源比赛数据，用脱敏数据形成开源数据集，推进产学研合作。”

无论如何，业界针对类似攻击已经开始从认知和技术上构建更高标准的应对措施。

邹皓向记者透露，目前信通院已经研究和开发相关的安全能力评估标准和工具，帮助技术提供方、技术使用方提升生物特征识别系统的安全性和可靠性。在他看来，国内机构应当在技术研发、标准化建设和国际合作上进一步加强投入，并进一步建立健全数据保护机制。

大模型打响“价格战” 垂直赛道或成破题点

本报记者 蒋牧云 何莎莎
上海 北京报道

近期，国内大模型市场“硝烟弥漫”，阿里云、百度智能云、腾讯、科大讯飞等厂商旗下的大模型费用纷纷下调。在采访中，多位业内人士告诉《中国经营报》记者，此番降价更多是各厂商之间在进行市场份额的争夺。

值得思考的是，“价格战”的背后，也反映出目前业内的大模型产品或服务同质化情况较为严重。多位业内人士指出，业内的大模型采用了类似技术架构和算法，并使用公共数据进行训练，激烈竞争下，急于推出产品，又使得厂商选择已获市场验证的服务模式，最终导致产品同质化现象严重。由此，各厂商如何在激烈竞争下寻求各自的突破口成为重要议题，其中，更具专业性的行业垂直大模型则显示出更大的发展潜力。

争夺市场份额

尽管较其他互联网大厂的模型发布时间较晚，但豆包大模型的价格却成功以企业级定价0.0008元/千Tokens（字符串），一经公布便成功“奇袭”。

不久后，多家大模型厂商也纷纷降价。阿里云宣布通义千问9款商业化及开源系列模型降价，其中，通义千问主力模型Qwen-Long，API输入价格从0.02元/千Tokens降至0.0005元/千Tokens，降幅达到97%；不久前发布的旗舰款大模型Qwen-max，API（数据服务接口）输入价格降至0.04元/千Tokens，降幅67%。百度智能云则在同一天迅速响应，宣布文心大模型两大面向企业的主力模型ERNIE Speed、ERNIE Lite全部免费。

隔天，腾讯云也公布全新大模型升级方案，主力模型之一混元-lite模型，API输入输出总长度计划从目前的4k升级到256k，价格从0.008元/千Tokens调整为全面免费。此外，多款混元大模型API输入价格都有50%—87.5%的降幅。同日，科大讯飞也宣布，讯飞星火API能力正式免费开放。其中，讯飞星火Lite API永久免费开放，讯飞星火Pro/Max API低至0.21元/万Tokens。

事实上，大模型产品自身拥有一定的降价逻辑，技术不断调整、迭代的情况下，推理成本、调用成本的下降也能不断优化大模型产品的价格。对比海外的OpenAI，同样有着不断降价的趋势。自2023年年初以来，OpenAI进行了4次产品降价，最近一次为5月13日发布的GPT-4o，价格下降了50%。

对于价格调整的考虑，阿里云相关负责人告诉记者，此次的降价得益于公共云的技术红利和规模效应。例如，阿里云基于自研的异构芯片互联、高性能网络HPN7.0、高性能存储CPFS等核心技术和产品，构建了极致弹性的AI算力调度系统，结合百炼分布式推理加速引擎，大幅压缩了模型推理成本，并加快了模型推理速度。

不久前百度在AI开发者大会上也提到，相比一年前，文心大模型的算法训练效率提升到了原来的5.1倍，周均训练有效率达到98.8%，推理性能提升了105倍，推理的成本降到了原来的1%。

不过，推理成本的降低并不代表研发成本的降低，尤其对比海外的芯片性能以及先发优势，国内大模型的成本优化和盈利能力还有很大的挖掘空间。比如，百度2023年财报显示，公司研发支出为242亿元，较2022年增长4%。百度在财报中表示，这主要由于支持生成式AI研发投入的服务器的折旧开支及服务器机架费增加所导致。

类似的，科大讯飞近期公布的一季报显示，报告期内，公司实现营业收入较去年同期增长26.27%；但第一季度归母净利润及扣非净利润分别较去年同期减少2.42亿元和1.02亿元。亏损增加的原因之一就是在通用人工智能认知大模型方面的投入。报告显示，公司2024年第一季度在大模型研发以及核心技术自主可控和产业链可控，以及大模型产业落地拓展等方面，新增投入约3亿元。其中，研发费用8.42亿元，较去年同期增加1.26亿元。

在投入巨大且持续上升的阶段，对于此番大模型费用密集降价，有大模型企业人士向记者直言：“实际还是为了抢占市场份额，由于有的厂商大幅降价，其他参与者也只能跟上。”对此，也有多位业内人士向记者表示，大模型的研发，尤其是通用大模型需要的算力、基础设施投入之大，更多还是大厂之间的比拼，市场的其他参与者很难与之竞争。

专业性服务不可替代

值得思考的是，我国大模型实际仍处于发展初期。在这一阶段就开始“价格战”，在业内看来是市场产品同质化严重所导致。北京市社会科学院研究员王鹏向记者表示，从当前市场上的大模型产品来看，很多都采用了相似的技术架构和算法，导致功能上的高度相似性。这种同质化现象使得厂商难以通过产品创新来吸引消费者，因此只能通过降低价格来争夺市场份额。

关于为何会产生同质化现象，王鹏表示，首先是技术成熟与算法相似性，导致了大模型在功能和性能上的趋同。其次，训练数据集的相似性也是原因之一，大模型的训练需要大量的数据，目前很多厂商都使用了相似的公开数据集进行训练，这也加剧了大模型之间的相似性。最后，在于市场竞争与用户需求，在激烈的市场竞争中，为了迅速占领市场，厂商更倾向于推出与市场上已有产品相似的大模型，因为这样可以降低研发成本和风险，同时可以满足用户对大模型的基本需求。

但不可否认的是，同质化背景下，降价对于业务的拓展有着一定效果。记者也观察到，不少专业性较强的企业，如金融机构近期都加强了与大模型厂商之间的合作。比如，泛华控股集团近日与百度智能云举行战略合作签约仪式，将共同打造AI保险销售助理“度晓保”，希望通过AI大模型使客户享受到更加专业、便捷的服务。此外，基金投顾业务试点机构盈米基金也接入通义千问，用于升级旗下基金投顾平台“且慢”的AI智能助理“小顾”，为用户提供更加智能、全面的投资服务。

那么，此次通用大模型的集体降价，是否也会挤压到垂直领域，如金融大模型的生存空间？大模型的降价潮是否会向垂直大模型蔓延？

王鹏表示，通用大模型降价后，原本考虑使用金融等垂直领域大模型的企业或个人，可能会因为价格因素而转向通用大模型，这将对垂直领域大模型的业务空间造成一定程度的挤压。与此同时，面对众多相似的大模型产品，客户在选择时可能会感到困惑和无所适从，这进一步加剧了市场竞争的激烈程度。因此，王鹏判断，随着通用大模型市场竞争的加剧，价格战有可能会向垂直领域蔓延。不过，垂直领域大模型由于其专业性和定制化需求，可能会在一定程度上抵御“价格战”的冲击。这是因为垂直领域大模型通常针对特定行业进行优化和定制，因此在某些专业领域仍具有不可替代性。

对此，也有金融机构人士告诉记者，目前机构对通用和金融垂直大模型均有过测试，从实际的使用效果来看，知识幻觉的情况还是较为明显。记者也了解到，目前大模型在金融机构的应用还是在营销素材或话术的生成，以及企业智能办公系统等方面。

由此，不论是通用大模型还是垂直大模型，如何突破同质化成为重要议题。在王鹏看来，从通用大模型的角度来看，需要持续提升通用性和易用性，使其能够适应更广泛的应用场景和用户需求。同时，通过不断增加新的功能特性来吸引用户，如多模态支持、跨语言处理等。从垂直大模型的角度来看，则需要深入行业需求，针对特定行业的需求进行深度定制和优化，提供专业化的解决方案。最为重要的是整合专业知识和数据资源，或与行业内的领先企业进行深度合作，通过针对性的专业知识和数据来提供更专业化的解决方案和服务。

综合看接下来的大模型竞争格局，中关村物联网产业联盟副秘书长袁帅向记者表示，在通用大模型市场，由于“价格战”的持续进行，市场竞争将更加激烈。厂商需要不断提升技术实力和服务质量，以应对市场竞争。在垂直领域大模型市场，其特定行业中的专业性和深度优势仍然无法替代。因此，垂直领域大模型仍将保持一定的市场规模和增长潜力。同时，随着技术的进步和市场的成熟，垂直领域大模型的成本也有望进一步降低，从而推动价格下降，这将使得垂直领域大模型在市场中更具竞争力。