服务“国之大者” 助力全面创新

为AI“假面”装上识别器

华政跨学科团队用法律与技术反击AI诈骗

2024年3月，一位老人接到“孙子”的求救电话，称与人发生口角、冲动伤人，需立即赔偿2万元救急。老人心急如焚，迅速取出现金交给对方。直到真正的孙子打来电话，他才发现受骗——那个熟悉的声音，竟是犯罪分子用AI技术伪造的。这类隐蔽性极强的AI诈骗案件，正在全国多地频繁发生。公安部数据显示，2025年一季度全国AI换脸诈骗案环比激增45%；国家反诈中心统计显示，今年AI诈骗案件同比增长超1900%。就在监管部门和科技企业为破解这一难题积极探索之际，来自华东政法大学的15名学生用三年时间构建起一个集“预警—鉴伪—普法”功能于一体的综合防治平台。2025年11月，该项目“生成式人工智能诈骗风险及其法治防控路径研究”获得第十九届“挑战杯”全国大学生课外学术科技作品竞赛特等奖。

青年报见习记者 王馨怡

▲华东政法大学团队获得第十九届“挑战杯”全国大学生课外学术科技作品竞赛特等奖。

◀ AI换脸技术案例分析。

本版均为受访者供图

从纸面到实践

2022年初，华东政法大学刑事法学院的程雅琳与7名法学专业同学组成项目组，最初的设想是围绕AI诈骗进行法律层面的调研和政策研究。

随着调研逐步深入，现实的残酷逐渐显现。他们通过上海市松江区公安分局反诈中心了解到，AI诈骗一直就在身边，已有不少高校教师曾收到AI合成的不雅照片勒索。受害者虽知照片是伪造的，但在技术上无法提供具有法律效力的鉴定报告，难以启动司法程序。同时，名誉勒索案件一旦报案，意味着将“不雅照片”公之于众，即使最终洗清冤屈，过程中的舆论压力也已构成二次伤害，许多人因此选择沉默与妥协。

调研中，一名女孩的电话让程雅琳至今难忘。听筒里，女孩的声音几乎破碎：“我知道那不是我，但我没法证明那不是我。我不敢告诉父母，不敢面对同学，我觉得自己的人生被彻底毁了……”

这些真实的困境让团队意识到，受害者不仅需要事后的法律定性，更迫切需要第一时间维护尊严、“自证清白”的技术工具。他们做出了一个大胆决定：跳出单纯的理论研究，开发一个法律与技术融合、真正可用的实战系统。

这一转型得到了项目指导老师、刑事法学院熊波教授的支持。他指出，面对AI诈骗这类新型风险，单纯依靠法律分析已显不足，他鼓励学生打破学科壁垒，通过课题研究揭示技术风险的隐蔽性，唤醒公众的主动防控意识。

从“不可能”到“可能”

“你们几个本科生，凭什么能解决连大厂都在攻坚的难题？”项目启动初期，来自业内专家的质疑直截了当。

但他们选择了坚持。2022年9月，程雅琳主动联系上了同校计算机专业的学生团队。当她真诚地展示大量真实案例和受害者访谈记录后，7名计算机专业学生深受触动，随即加入。由此，一个由15人组成的跨学科团队，因一份共同的责任感而凝聚。

然而，跨学科合作之初，法学生与计算机学生之间经历了激烈的思维碰撞。熊波称之为“结构性的、贯穿始终的挑战”。法学背景的学生习惯规范性思维，讨论公平、治理、责任；计算机背景的学生则擅长工程化思维，关注算法、模型、数据流，目标逻辑一度存在偏差。

为打破壁垒，熊波充当“翻译”和“桥梁”。他组织跨学科研讨，让法学组讲法律框架，让技术组讲AI原理，并通过“耦合性任务”围绕“AI生成内容标识”共建共研。经过半年的磨合后，团队才真正实现融合。

如果仅由计算机专业的学生独立完成，这个项目可能会大不相同。法学背景的参与带来了关键的“化学反应”。

最终产出的“电子数据取证监督机器人”，不仅能通过面部光影、微表情等识别伪造痕迹，更能自动提取视频元数据。使用流程简洁直观：受害者打开系统，将可疑视频复制粘贴到平台网页，点击分析按钮，系统即开始对视频真伪性进行全面检测。分析完成后，平台输出直观的可视化图表，显示“伪造”或“真实”的判定结果。

“目前系统还不能直接生成具有法律效力的权威鉴定报告，”程雅琳说，“这需要专家或权威机构的背书。我们计划在平台正式移交公安机关后，配合相关部门开发这一功能，确保报告能真正用于司法程序。”

2024年3月，该系统成功获批国家专利。

构建“三位一体”法治体系

除了研发出能一键鉴伪的“电子数据取证监督机器人”，项目组在三年间还进行了大量实证研究。他们跨越3000多公里，对23个省市进行了深入走访，收集了5000个案例和1680份有效问卷。

这项一线调研揭示了AI诈骗背后更深层的制度漏洞和治理困境。因此，项目组构建了一套涵盖立法、司法和执法全链条的“三位一体”防治体系，旨在厘清平台与数据源的法律责任，破解权责模糊难题。

在立法层面，针对“源头难溯”的问题，他们建议所有公开传播的AI生成内容必须嵌入双层数字水印，就像给AI生成内容打上“数字身份证”。

在司法层面，面对高科技证据，法官往往面临“看不懂代码”的技术壁垒。针对“鉴定难判”的问题，团队建议设立“技术调查官”制度，让懂技术的人辅助法官进行事实认定。这一制度类似知识产权案件中的“技术专家陪审”，能有效弥补司法人员在AI技术鉴定方面的专业短板。

在执法层面，传统反诈往往是事后追查。团队建议建立“AI诈骗风险联防联控中枢”，推行“红橙黄”三级动态预警。这意味着将反诈端口前移，在诈骗发生前或进行中就进行分级阻断，而不是等受害者遭受损失后再介入。

这些研究成果并非停留在理论层面。项目组通过智库专报等多种渠道，将建议传递至中央及省级决策部门。部分思路和建议被认为具有参考价值，已被纳入相关部门对《生成式人工智能服务管理暂行办法》的审查意见中。

尚未完成的答卷

尽管项目获得特等奖、技术获得专利，程雅琳仍保持清醒的谦逊：“我们只是迈出了第一步。”目前平台仍处于测试阶段，要真正推广应用，还需打通数据对接通道，完成审批和安全认证，在真实场景中不断试点优化。

三年的项目历程，让团队有了新的认知。程雅琳坦言，随着AI技术不断迭代，检测模型也需要持续升级优化，“但最重要的不是我们做出了多完美的系统，而是用青年视角发现了被忽视的问题，证明了法学与计算机结合这条路径的可行性。”

正因身处青年学生这一独特位置，他们更能看到那些被主流忽略的痛点——高知群体的“名誉勒索”困境，偏远地区青少年的反诈教育空白，以及受害者“无法快速证明清白”的无力感。“如果这套系统未来能成功接入公安、司法系统，哪怕只是减少一个具体的人受到伤害，那都是值得的。”程雅琳说。

熊波教授对此深表认同。他希望这一项目不仅成为一个反诈工具，更能从“项目”走向“范式”，推动法律与技术协同治理机制的探索。

三年的项目历程，让团队有了新的认知。他们意识到，虽然要将一个学术原型转化为解决社会问题的利器仍需时日，但这份研究的价值已然显现。“我们用三年时间，证明了法律的严谨与技术的锋芒可以同行。”

这是一份尚未完成的答卷，但答题的方向已经明晰。