2026 网络安全宣传作品 · 互动作品

镜 鉴

AI 换脸诈骗拆解实验室

眼见可以被伪造，核验才是防线

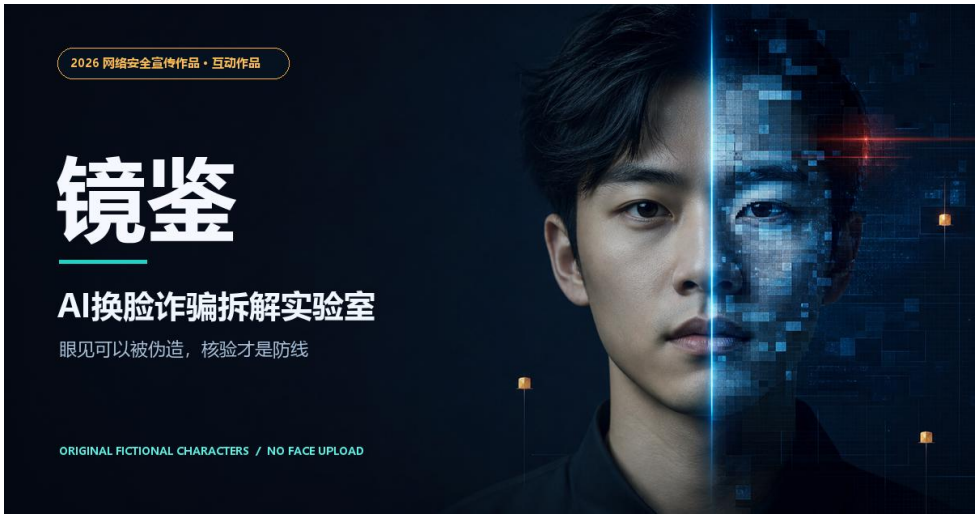

图 1 《镜鉴》原创视觉封面

作品定位 面向高校师生的网络安全公益互动 H5。作品以虚构熟人视频诈骗为情境，通过剧情选择、换脸过程科普、差异热力图和反诈决策训练，引导体验者形成“停、核、护、报”的身份核验行动。

申报信息	内容
作品类型	互动作品（H5）
报送单位	信息与通信工程学院
作者	苗桐郡、罗天响
联系电话	15832333564
正式网址	https://kmyc-spirit-2026-d5ea11ad8a1a951-1439949158.tcloudbaseapp.com/

当“眼见”不再可靠，防线应从识别转向核验

AI 换脸、语音克隆等生成合成技术降低了身份冒充的门槛。诈骗者往往把逼真的声画与“紧急、保密、转账、授权”等高压话术叠加，使受害者在来不及思考和交叉验证的情况下作出决定。对普通师生而言，仅靠寻找画面破绽既不稳定，也难以迁移到新的伪造方式。

《镜鉴》因此不把重点放在“猜出哪张脸是假的”，而是训练一套在真假难辨时仍然有效的安全动作：先退出对方制造的节奏，再通过独立渠道核验身份，保护验证码、屏幕和账户权限，发生风险后立即止付、留证和报告。

三项学习目标

学习目标	体验结束后能够做到
识别风险组合	对“身份 + 紧急 + 金钱/权限”同时出现保持警惕
建立核验链	挂断原通话，改用通讯录、官方入口或线下渠道独立确认
完成应急处置	停止共享与转账，保护账户，联系银行、平台和公安机关

目标受众与应用场景

作品主要面向高校学生、教师和行政人员，可用于网络安全宣传周、课堂导入、主题班会、新生教育和校园新媒体传播。核心互动约 90 秒，完整体验约 3 分钟，无需登录或安装应用，手机和电脑均可直接访问。

作品边界 《镜鉴》不是换脸工具，也不是通用鉴伪器。它是一套帮助普通人在真假难辨时做出正确选择的反诈行为训练。

02 · 互动体验流程

让体验者在一次虚构来电中完成完整核验

阶段	体验内容	形成的认识
1 · 来电	虚构的“陈老师”发起紧急视频通话并提出转账要求	逼真声画不能单独证明身份
2 · 选择	体验者决定继续配合，还是挂断后独立核验	先切断对方控制的沟通节奏
3 · 鉴真	通过图像对比和 36 秒科普视频理解换脸过程与差异热力图	画面异常只能提示风险，不能代替身份核验
4 · 决策	完成共同秘密、官方渠道、屏幕共享和事后处置四轮选择	把反诈知识转化为可执行动作
5 · 行动卡	根据选择生成“停、核、护、报”反诈行动卡	带走一套可复用的核验方法

图 2 公开互动 H5 首屏（桌面端）

叙事策略 作品不展示真实转账页面，也不让体验者提交任何个人信息；通过可重来的虚构情境，让错误选择变成低风险、可解释的学习机会。

停、核、护、报：把知识变成顺序动作

记忆点	关键动作	要纠正的误区
停	挂断高压通话，不点击、不共享、不转账	继续留在对方通道里“再确认一下”
核	使用原通讯录、官方入口或线下渠道重新联系	让对方在当前视频中“证明自己”
护	保护验证码、屏幕内容、账户和付款权限	以为遮住部分通知就能安全共享屏幕
报	联系银行止付，保留证据，及时向平台和公安机关报告	删除记录或等待对方主动退款

鉴真台：讲清视觉线索的适用边界

鉴真台展示的是同一组已知原图与合成图之间的像素差异，用于帮助体验者理解合成内容可能留下局部变化。作品同时明确说明：压缩、弱网、光照和美颜都可能制造类似现象，差异热图不是通用检测结论，更不能作为转账或授权依据。

决策反馈：解释原因，而不是只判对错

每轮选择都会立即说明其后果。安全选项解释为什么有效；风险选项则揭示诈骗者如何利用小额试转、公开信息、群聊仿冒和持续套话扩大损失。评分服务于复盘，不排名、不收集成绩，也不对体验者作负面评价。面对语音克隆、盗号聊天、伪造客服或屏幕共享诈骗，同一套“停、核、护、报”仍然成立。

教育结论 可疑画面应当触发独立核验，而不是触发“我已经看出真假”的盲目自信。

把抽象技术风险，变成公众看得懂的过程

图 4 36 秒深度伪造科普视频关键画面

创新点	设计表达
从技术奇观转向行为训练	不以“换脸有多逼真”作为终点，而以“面对逼真伪造怎样行动”作为核心
把换脸过程讲明白	用源身份、目标画面、替换过程和合成结果四个步骤解释基本原理
把热力图边界讲清楚	展示本例高变化区域，同时明确热图不能判断任意视频真假
从看破绽转向完成核验	视频结尾直接给出挂断、换渠道、找本人和走流程的行动顺序
低门槛传播	纯前端 H5，无需登录和安装，适配手机与电脑，便于课堂和宣传活动复用

视觉与交互语言

整体采用“鉴证实验室”视觉风格：青色代表安全行动，琥珀色提示风险，红色用于合成标识和高风险警示。导航、按钮和反馈均使用明确文字，不依赖颜色单独传达状态；科普视频配有中文字幕，主要触控区域按移动端操作习惯设计。

公开作品只做教育，不处理体验者人脸

安全边界	公开 H5 的处理方式
身份与账户	无需注册、登录或填写姓名，不建立用户画像
照片与摄像头	没有上传控件，不调用摄像头或麦克风
人脸处理	没有人脸处理后台，不向浏览器提供换脸接口
成绩与行为数据	行动卡在本地生成，不保存选择、分数或设备信息
科普内容	仅播放持续带有 AI 生成标识的虚构人物录像，讲解换脸过程、热力图与核验方法

本地演示舱的安全约束

真实算法能力仅用于现场非商业教育演示：源身份固定为不存在的虚构人物；启用摄像头前需明确同意；画面仅在内存中处理，不录制、不上传、不导出；每帧持续显示 AI 生成标识；单次体验最长 60 秒，并提供一键急停和异常回退录像。

原创素材与权利边界

剧情人物和演示身份均为项目组面向本作品定向生成并二次设计的虚构角色，不指向、不模仿任何真实人物或公众人物，也不使用网络下载肖像。素材来源、生成方式和使用目的均留档。开源组件按相应许可使用；本地模型权重不进入网站、源码包或报送附件。

法规与标准依据

作品参照《互联网信息服务深度合成管理规定》《人工智能生成合成内容标识办法》和 GB 45438—2025《网络安全技术 人工智能生成合成内容标识方法》，在合成图片、录像和演示界面中持续呈现显式标识，并明确反对删除、弱化或隐藏标识。

网站以通俗语言解释三项公众须知：生成合成内容应依法标识；未经同意不仿冒、不传播；遇到疑似换脸诈骗应保留证据并及时联系平台、银行和公安机关。

一次体验，带走一套可复用的核验方法

《镜鉴》把深度伪造这一较强技术话题转化为普通师生可以理解、参与和复述的行动语言。体验者无需掌握算法原理，也能在几分钟内经历风险、作出选择、获得解释并形成行动卡，适合在校园网络安全教育中低成本、常态化使用。

可推广的应用方式

场景	使用方式
宣传周与展台	扫码完成互动，配合大屏播放深度伪造科普视频
课堂与班会	以虚构来电作为导入，围绕错误选项组织讨论
新生与教职工培训	将“停、核、护、报”作为身份核验基础流程
校园新媒体	通过国内 HTTPS 链接和二维码直接传播，无需安装

正式访问

国内 HTTPS 地址 https://kmyc-spirit-2026-d5ea11ad8a1a951-1439949158.tcloudbaseapp.com/

扫码进入《镜鉴》互动作品

访问提示：当前使用腾讯云 CloudBase 国内默认域名，首次打开可能出现腾讯云官方安全提醒，点击“确定访问”后即可进入作品。

最终记忆点 眼见可以被伪造，核验才是防线。面对“身份 + 紧急 + 金钱/权限”：停、核、护、报。