ICS 35.040
CCS L80

团　　体　　标　　准

T/ISC XXXX—XXXX

智能身份伪造防范认证规范

Specification for Preventing Intelligent Identity Forgery in Authentication

（征求意见稿）

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

中　国　互　联　网　协　会　　发 布

目　　次

前　言

本文件按照GB/T 1.1—2020《标准化工作导则 第1部分 标准化文件的结构和起草规则》的规定起草。

请注意本文件的某些内容可能涉及专利。本文件的发布机构不承担识别这些专利的责任。

本文件由中国互联网协会提出并归口。

本文件起草单位：中国信息通信研究院、蚂蚁科技集团股份有限公司、浙江大学、四川大学、复旦大学、北京智慧易科技有限公司、原力金智（北京）科技有限公司、汇海银河（北京）科技有限公司、中电金信软件有限公司、数字广东网络建设有限公司

本文件主要起草人：吴寒冰、王景尧、吴荻、王维强、刘健、刘菁菁、黄余格、王振明、唐佳伟、赵佳妮、周晟、周吉喆、李晟、关涛、冯超、李旭刚、雷震、詹港傲、吕逸良、李祖金。

引　言

随着生成式人工智能（AIGC）与深度合成技术的飞速发展，智能身份伪造已成为威胁金融与政务领域安全的关键风险。近年来，利用深度伪造（Deepfake）技术伪造人脸视频进行远程开户、利用声音克隆技术冒充企业高管实施诈骗，以及通过注入攻击绕过活体检测等典型安全事件频发，严重损害了用户财产安全与公共服务秩序。

现行的 GB/T 41819、GB/T 38671 及 T/IIFAA 3001.1 等相关标准，主要侧重于单一模态的身份鉴别，尚未有效覆盖图、视、音多模态联合认证，缺乏针对注入攻击的有效防御机制以及 AIGC 生成内容的标识规范，难以应对当前复杂多变的伪造威胁。

本标准规定了多模态联合认证的技术要求、注入攻击防御策略及 AIGC 内容标识方法。本标准适用于金融、政务及其他高安全需求场景下的身份认证系统设计、开发与检测，旨在提升系统对新型智能伪造攻击的防御能力，保障数字身份的真实性与安全性。

智能身份伪造防范认证规范

1 范围

本文件规定了图视音多模态身份安全认证的术语定义、总体原则、技术要求、安全要求、认证流程、测试评估、管理要求，以及高风险金融与政务场景专项认证细则。

本文件适用于金融、政务、通信、互联网、安防等涉及身份核验的各类场景，为智能身份伪造防范、身份认证系统设计、开发、部署、测评及监管提供技术依据，同时适用于相关产品的安全认证与合规核查，其中高风险专项细则强制适用于金融、政务领域高敏感身份核验业务。

本文件不适用于：

（a）仅基于密码、短信验证码等非生物特征的单一因子认证；

（b）公安刑事侦查、司法鉴定等专业取证场景；

（c）军事、保密通信等特殊领域的身份认证系统。

2 规范性引用文件

下列文件中的内容通过文中的规范性引用而构成本文件必不可少的条款。其中，注日期的引用文件，仅该日期对应的版本适用于本文件；不注日期的引用文件，其最新版本（包括所有的修改单）适用于本文件。

GB 45438-2025 网络安全技术 人工智能生成合成内容标识方法

GB/T 25069-2022 信息安全技术 术语

GB/T 22239-2019 信息安全技术 网络安全等级保护基本要求

GB/T 35273-2020 信息安全技术 个人信息安全规范

GB/T 41819-2022 信息安全技术 人脸识别数据安全要求

GB/T 38671-2020 信息安全技术 远程人脸识别系统技术要求

GB/T 41807-2022 信息安全技术 声纹识别数据安全要求

GB/T 42981-2023 信息技术 生物特征识别 人脸识别系统测试方法

GB/T 44248-2024 信息技术 生物特征识别 人脸识别系统应用要求

GB/T 41772-2022 信息技术 生物特征识别 人脸识别系统技术要求

GB/T 37078-2022 信息安全技术 生物特征识别安全技术要求

GA/T 1093-2021 人脸识别身份验证技术要求

GA/T 1093-2023 出入口控制人脸识别系统技术要求

JR/T 0197-2020 金融数据安全 数据安全分级指南

JR/T 0185-2020 商业银行应用程序接口安全管理规范

YD/T 4092-2022 电信和互联网行业生物特征识别安全要求

YD/T 3830-2021 多模态生物特征识别系统安全技术要求

DB11/T 2252-2024 信息安全 人脸识别防对抗样本攻击测试要求

T/IIFAA 3001.1-2021 远程人脸识别应用技术规范 第 1 部分：金融账户管理

T/ZFIDA 0003-2024 金融应用 虚假数字人脸检测

T/CITIF 006-2023 远程人脸识别系统防注入攻击技术要求及测试方法

ISO/IEC 30107-1：信息技术 生物特征识别呈现攻击检测 第 1 部分：框架

ISO/IEC 30107-2：信息技术 生物特征识别呈现攻击检测 第 2 部分：数据格式

ISO/IEC 30107-3：信息技术 生物特征识别呈现攻击检测 第 3 部分：测试与报告

3 术语和定义

下列术语和定义适用于本文件。

3.1 智能身份伪造

利用人工智能、深度学习、图像/音频/视频合成、篡改等技术，对自然人的生物特征（如人脸、声纹）、身份凭证（如身份证件、生物特征采样数据）或行为表现（如动作姿态、言谈逻辑）进行仿真、置换、篡改或重构，从而创造出能够通过身份核验系统、足以产生社会危害或欺诈后果的虚假身份标识的过程。包括但不限于制作高精度人脸面具、Deepfake 人脸伪造、AI 合成语音、声音克隆、篡改证照图像等。

3.2 图视音多模态身份认证

基于图像（静态人脸、证照、生物特征图像）、视频（动态人脸、行为动作）、音频（声纹、语音内容）三种及以上模态数据，通过多特征融合、交叉核验，实现自然人身份真实性、合法性验证的技术方式。

3.3 跨模态一致性校验

对图、视、音多模态数据进行语义、时序、物理特征、生物特征的匹配性验证，判断各模态数据是否来源于同一主体、是否存在伪造篡改痕迹的技术过程。

注：匹配性验证包括但不限于语音内容与唇形动作的时序一致性；证件材质、纹理与环境光影的物

理一致性；人脸生物特征在图像、视频模态下的一致性等。

3.4 身份伪造检测

通过特征分析、算法核验、规律校验等手段，识别身份数据、生物特征、交互行为中的伪造、篡改、冒用风险，判定身份认证对象真实性的技术方法。

3.5 高风险金融/政务场景

指涉及资金划转、资产变更、敏感隐私信息查询、法定权限授予、电子签章、证照签发等高敏感、高安全等级的金融与政务业务场景，尤其是政务和金融结合的场景，比如公积金账户信息变更及公积金提取。

3.6 攻击误识率 APCER

指伪造攻击样本被错误判定为真实样本的比例，数值越低代表抗伪造攻击能力越强。

计算公式为：APCER = （错误接受的攻击样本数 / 总攻击样本数） × 100%。

3.7 真实误拒率 BPCER

指真实样本被错误判定为伪造样本的比例，数值越低代表认证通过率与准确性越高。

计算公式为：BPCER = （错误拒绝的真实样本数 / 总真实样本数） × 100%。

3.8 比对误识率 FAR

指非本人被认作本人的比例，数值越低代表可用性与准确性越高。

3.9 比对拒识率 FRR

指本人被认作非本人的比例，数值越低代表可用性与准确性越高。

3.10 安全采集 Secure Capture

通过设备完整性证明、受控采集管线、采集时密码学签名等机制，确保生物特征数据在采集环节真实可信、防注入、防篡改的技术过程。

3.11 设备完整性证明 Device Attestation

基于硬件信任根，向认证服务方证明采集终端的启动链、操作系统、采集应用未被篡改的密码学机制。

3.12 受控采集管线 Controlled Capture Pipeline

从图像传感器、麦克风到采集应用之间的数据通路具备来源真实性校验能力，可识别虚拟摄像头、音频回环、Hook 注入等伪造源的采集机制。

3.13 呈现攻击工具种类 PAI Species

指攻击者在呈现攻击过程中使用的物理或数字媒介的集合。根据攻击手段的物理属性和生成方式，可分为：传统 PAI： 物理介质类，如打印照片、显示屏、硅胶面具、录音设备。生成式 PAI： 由 AIGC 技术产生的伪造介质，如 Deepfake 人脸视频、AI 合成音频。

3.14 数字注入攻击 Digital Injection Attack

通过技术手段，向系统的数据流中注入虚假、合成或恶意的媒体内容（如音频、视频流），以欺骗身份验证系统或在线交互对象的网络攻击方式。

3.15 注入攻击呈现分类错误率 IAPCER

指针对数字注入攻击的特定分类错误率。用于衡量系统对绕过物理传感器的数字流攻击的检出能力。

计算公式为：IAPCER = （错误接受的注入攻击样本数 / 总注入攻击样本数） × 100%。

3.16 系统级评估指标 IAPMR

指在全系统测试中，攻击者通过伪造呈现（包括数字注入）被系统比对模块判定为合法目标的比例，反映了防御系统在“检测+比对”双重防线下的整体脆弱性。

3.17 深度伪造 Deepfake

利用人工智能（特别是深度学习算法）来合成或篡改音视频内容的技术。

3.18 AIGC 数字人/虚拟形象

通过生成式 AI 技术创建的、具有完整外貌、语音和交互能力的数字实体。与 Deepfake 依赖特定真实人物素材不同，此类形象可能完全脱离物理现实存在，具备虚构的身份属性，常被用于冒充客服、业务员等身份实施欺诈。

3.19 声音克隆 Voice Cloning

通过少量目标音频样本训练个性化声码器，精准还原目标人物的音色、韵律和说话习惯，产生的语音内容在声纹特征层面与本人高度相似。

3.20 声纹 Voiceprint

对语音中所蕴含的、能表征和标识说话人的语音特征，以及基于这些特征（参数）所建立的语音模型的总称。

3.21 对抗样本攻击 Adversarial Example Attack

通过向原始输入数据（如图像、语音、文本）中故意添加人类难以察觉的细微扰动，生成“对抗样本”，以欺骗人工智能模型并使其产生错误预测或分类结果的一种攻击方式

4 缩略语

TP　True Positive　真阳（真实样本判定为真实）
TN　True Negative　真阴（伪造样本判定为伪造）
FP　False Positive　假阳（伪造样本判定为真实）
FN　False Negative　假阴（真实样本判定为伪造）
Acc　Accuracy　准确率
TPR　True Positive Rate　召回率
FPR　False Positive Rate　误检率
APCER　Attack Presentation Classification Error Rate　攻击误识率
BPCER　Bonafide Presentation Classification Error Rate　真实误拒率

IAPCER　Injection Attack Presentation Classification Error Rate　注入攻击呈现分类错误率（针对数字注入）

IAPMR　Impostor Attack Presentation Match Rate　冒充者攻击呈现匹配率（系统级测试指标）

FAR　False Acceptance Rate　比对误识率

FRR　False Rejection Rate　比对拒识率

EER　Equal Error Rate　等错误率（当FRR = FAR或BPCER = APCER时的数值）

5　总体原则

5.1　安全优先原则

身份认证全流程以防范伪造风险为核心，优先保障身份核验准确性、安全性，兼顾认证效率与用户体验，高风险场景执行最高安全等级防护。

5.2　分级适配原则

按照业务风险程度划分安全等级，实施差异化认证防护策略，低风险场景兼顾便捷，高风险场景严控安全。

5.3　多维度核验原则

单一模态认证不可作为最终核验依据，必须采用多模态交叉验证、多算法融合校验，降低伪造攻击成功率，高风险场景强化核验维度。

5.4　合规可控原则

严格遵循个人信息保护、数据安全、生物特征管理、金融监管、政务信息化相关法律法规，身份数据采集、存储、使用全程合规，权限可控、可追溯、可解释、可核查。

5.5　动态适配原则

针对新型智能伪造技术，持续更新认证算法、检测模型、安全策略，具备对抗迭代式伪造攻击的能力，定期开展高风险场景安全升级。

5.6　最小权限原则

身份认证过程仅采集必要信息，严格限定认证数据使用范围，杜绝超范围采集与滥用，高风险场景执行权限最小化、操作可审计。

6　分级安全认证要求

本文件按业务风险等级将身份认证划分为L1 低风险、L2 中风险、L3 高风险三个等级。各等级适用场景、模态要求、防护措施、交互方式、安全指标需满足表 1 要求，其中 L3 等级需同步满足本文件中金融/政务高风险场景专项核验要求。

表 1 风险等级划分及管控要求表

安全等级	适用场景	物理防护要求	数字防护要求	活体认证安全指标	比对指标	系统级指标
L1 低风险	普通注册、社交登录	防御静态照片、低清屏幕	无强制要求	APCER≤5% IAPCER≤5% BPCER≤5%	FAR≤10^{-4} FRR≤3%	≤5%
L2 中风险	个人信息修改、普通政务	防御高清屏幕重放、2D 面具	基础注入防御，基础 Deepfake/AIGC 防御	APCER≤1% IAPCER≤1% BPCER≤5%	FAR≤10^{-5} FRR≤2%	≤2%
L3 高风险	金融交易、敏感交易	防 3D 面具、翻拍，物理环境安全校验	专业级注入防御，防御实时 Deepfake/AIGC 等	APCER≤0.1% IAPCER≤0.1% BPCER≤10%	FAR≤10^{-6} FRR≤2%	≤1%

7 系统输入/输出信息

7.1 系统输入信息

系统输入信息包括：

a） 生物特征流：高清影像流（RGB/IR/Depth）、高清音频流（采样率≥16kHz）、证件原件影像；

b） 环境特征：设备指纹（设备 ID、传感器状态）、网络环境（IP 地理位置、VPN 检测）、操作行为特征（传感器轨迹、点击节奏）；

c） 交互指令：系统下发的动态挑战码（随机数字、随机文本、指定动作序列）。

7.2 系统输出信息

系统输出信息格式：

a） 输出结果宜采用JSON格式文件；

b） 输出文档应包含认证结果、认证原因等信息，输出文件参考表2。

表2 输出形式参考样例

JSON 文档	说明
{"request_id": "req_20260519_abc123",	请求 id
"status": "success",	状态
"result": {	
"decision": "pass",	认证结果，pass/reject/manual_review
"overall_score": 0.98,	认证总分数
"confidence_level": "high",	置信等级
"authentication_scores": {	

"physical_anti_forgery_score": 0.001,	物理防伪分
"digital_anti_forgery_score": 0.011,	数字防伪分
"interactive_action_score": 0.97,	交互动作分
"verification_score": 0.8},	比对分
"timestamp": "2024-10-27T10:00:00Z"}	时间戳

8 图视音多模态身份安全认证规范

8.1 基础要求

认证系统需兼容 L1-L3 全等级认证模式，可根据业务场景自动适配对应安全等级核验要求。

L1 等级可采用单模态或单因素认证（如静态密码、短信验证码），但不得将单一的、可被轻易重放的静态密码作为高风险操作的认证依据。L2 等级应执行双模态认证，L3 等级应执行三模态联合核验，当单模态防伪能力达到 L2 或 L3 的认证安全指标时，可经评估后作为双/三模态组合的并列选项。

身份数据采集需满足真实性、完整性、时效性要求，禁止使用过期、篡改、非实时数据开展认证。

8.1.1 安全采集要求

本节作为 8.2、8.3、8.4 各模态采集要求的共同前置，规范图像、视频、音频采集过程的真实性、完整性、可信性。采集终端应符合下述安全设计准则，并满足附录 B 中规定的强制性指标。

a） 采集环境真实性

1） 采集终端应通过设备完整性证明机制向认证服务方证明运行环境可信。硬件信任根（TEE、SE、StrongBox 或国产可信执行环境）可用时应优先使用。

2） 移动端应用在不具备硬件信任根的情况下，应至少完成软件层完整性校验，包括应用签名核验、Root 或越狱检测、调试器检测、Hook 检测。

3） 采集应用应通过代码签名校验确认与备案版本一致，禁止重打包、注入修改后的应用执行采集任务。

b） 受控采集管线

1） 采集应通过系统原生底层接口直接获取传感器数据，禁止经由可被运行时替换的中间层。

2） 应具备虚拟摄像头、虚拟麦克风、音频回环、模拟器、跨进程帧投递的识别能力，识别到上述情形应直接终止采集并判定认证失败。

3） 应同步采集传感器辅助信号，包括陀螺仪、加速度计、环境光、网络环境、地理位置等可获取项，作为采集真实性旁证。

c） 采集时签名

1） 应在采集会话结束时，对采集数据整体哈希、采集时间戳、设备标识、辅助信号摘要做一次数字签名。

2） 签名算法应支持 SM2、SM3 国密算法套件。采用国际算法时，应使用 ECDSA P-256 及以上、SHA-256 及以上。

3） 时间戳应来源于设备外部可信时间源。L3 等级应使用经授权的时间戳服务（TSA）。

d） 分级实施要求

L1 等级：完整性证明不强制；受控管线应具备虚拟摄像头识别能力；采集时签名可选；时间戳采用设备本地时间。

L2 等级：软件完整性校验强制，硬件证明可用时应启用；受控管线强制具备；采集时签名强制；时间戳采用可信外部时间源。

L3 等级：可控终端（金融/政务自助终端、专用核验终端）应强制启用基于硬件信任根的设备完整性证明；移动端在硬件信任根可用时应强制启用该证明机制，不可用时应通过服务端多因子核验等补充措施进行风险补偿。认证系统应强制具备受控采集管线能力，并应识别注入攻击。应强制执行采集时签名，签名密钥的保护应符合“9.5 密钥安全管理要求”的要求。时间戳应来源于具备国家授时资质的时间戳服务机构（TSA）。

8.2 图像身份认证安全规范

8.2.1 采集要求

a） 采集设备需具备防翻拍、防截图检测能力，支持图像质量判定（清晰度、光照度、完整性、无遮挡）。

b） 人脸图像需采集正面免冠、无遮挡、实时动态画面，禁止使用静态照片、打印图像、电子截图替代。

c） 证照图像需完整采集证照全部区域，支持证照防伪特征、篡改痕迹检测，识别 PS、拼接、合成等伪造操作。

8.2.2 伪造检测要求

a） 具备图像合成痕迹检测能力，识别像素异常、纹理失真、光照矛盾、几何畸变等 AI 生成/篡改特征。

b） 支持证照防伪校验，包括水印、二维码、防伪码、纸质纹理、印章真实性核验。

c） 对人脸图像，需完成活体特征初步判定，排除 2D 静态图像、面具、模型等伪造载体。

8.2.3 认证要求

a） 图像特征提取需采用合规算法，特征数据加密存储，禁止明文传输原始图像信息。

b） 图像比对准确率需匹配对应安全等级指标。

8.3 视频身份认证安全规范

8.3.1 采集要求

a） 交互时长：采集有效视频流时长应满足：L2 级不少于 3 秒，L3 级不少于 4 秒。

b） 数据完整性：视频流应保持连续，不得出现非物理因素导致的帧丢失、断流或图像冻结。采集数据应包含原始采集时间戳、终端设备唯一标识。

c） 质量保障：系统应具备实时视频质量评估能力，对图像模糊度、曝光水平、遮挡程度及卡顿率进行筛查，剔除画质不达标的无效样本。

8.3.2 伪造检测要求

a） 应检测视频深度伪造痕迹，包括但不限于面部微表情分析、帧间生理信号（心率、血流波动）异常、动作连贯性缺失、面部边缘模糊、光影失真中的一项或多项。

b） 应核验视频实时性，L3 等级通过随机动作指令、时间戳校验、场景变化判定，排除预录制伪造视频。

c） 应支持视频内容完整性校验，识别剪辑、拼接、换脸、AI 生成等篡改操作。

8.3.3 认证要求

a） 动态人脸特征与静态证照图像完成一致性比对，验证主体同一性。

b） L3 高风险场景需叠加唇语与语音同步性校验，提升伪造对抗能力。

8.4 音频身份认证安全规范

8.4.1 采集要求

a） L2 等级可采用固定语音采集，L3 等级必须采集实时随机语音数据，排除预录制音频。

b） 采集环境需降噪处理，识别静音、杂音、合成音频等异常状态。

8.4.2 伪造检测要求

a） 检测 AI 合成语音、语音转换、声音克隆痕迹，识别频谱异常、基频波动、节奏不连贯、生理发声特征缺失。

b） L3 等级校验语音实时性，通过随机指令响应、语音动态变化判定，排除预录制、循环播放音频。

c） 支持声纹唯一性核验，区分目标主体与伪造语音的特征差异。

8.4.3 认证要求

a） 声纹特征加密提取、加密存储，严禁原始语音数据泄露。

b） L3 等级音频认证需与视频、图像模态联动，完成语音内容、唇形、身份信息的一致性匹配。

8.5 跨模态联合认证规范

a） 应建立多模态数据对齐机制，校验图像、视频、音频数据的主体同一性、时序一致性、语义一致性。

b） 应采用多算法融合决策，单一模态核验不通过则整体认证失败，多模态结果加权判定认证可信度。

c） L3 高风险场景需叠加活体检测、设备环境检测、用户行为校验，形成“生物特征+行为特征+环境特征”三重核验。

d） 对跨模态不一致数据，直接判定为伪造风险，启动二次认证或人工复核。

9 通用安全技术要求

9.1 数据安全要求

a） 图视音身份数据采集遵循“最小必要”原则，仅采集认证必需信息。

b） 身份数据传输采用 SSL/TLS 等加密协议，存储采用加密算法，禁止明文存储、传输。

c） 认证完成后，非必要原始身份数据及时删除，留存数据需脱敏处理。

d） 严禁非法共享、泄露、篡改身份认证数据，保障数据完整性与保密性。

9.2 算法安全要求

a） 身份认证、伪造检测算法需通过安全测评，无后门、无漏洞，具备抗攻击能力。

b） 身份核验模型应具备识别并抵御对抗样本攻击的能力。系统在预处理阶段应包含降噪、去扰动、图像压缩或随机化处理，以降低输入样本中人为添加微小扰动导致的模型分类偏误。

c） 多模态融合算法需具备容错性，避免单一算法故障导致整体认证失效。

d） 身份认证与伪造检测核心算法模型应建立动态更新机制。运营方应至少每季度对模型抵抗新型伪造攻击的能力进行一次评估，并至少每半年根据技术发展和威胁形势完成一次必要的模型迭代升级，适配新型智能伪造技术，满足各等级安全指标要求。

9.3 系统安全要求

a） 身份认证系统应具备防攻击、防篡改、防渗透能力，抵御网络攻击、恶意破解。

b） 应建立异常预警机制，对批量伪造认证、高频攻击、异常核验等行为实时告警。

c） 系统应具备容灾备份能力，保障认证服务连续性，避免服务中断导致认证失效。

9.4 隐私保护要求

a） 单独同意机制：生物识别信息处理须取得用户单独同意，不得与其他个人信息处理同意合并；

b） 个人信息保护影响评估（PIA）事前要求：处理生物识别信息前应开展并留存评估报告；

c） 跨境提供评估义务：通过国家网信部门安全评估或个人信息保护认证；

d） 非唯一性原则：同一目的存在其他验证方式时，不得将人脸识别作为唯一验证方式，应提供非生物特征的备选认证通道；

e） 未成年人保护：对未成年人生物特征采集应取得监护人同意并适用更严格的存储与删除规则；

f） 用户撤回机制：支持用户注销身份信息、撤销授权，注销后彻底删除相关身份认证数据。

9.5 密钥安全管理要求

a） 硬件信任根优先

应具备硬件密钥库能力的实施环境，认证密钥应在硬件密钥库（HSM、TEE、SE、StrongBox 或国产可信执行环境）内生成、存储、使用，密钥全生命周期不应离开可信边界。L3 等级在硬件信任根可用时应强制使用。

b） 软件层密钥防护

1） 在不具备硬件安全模块的实施环境中，应通过代码混淆、运行时完整性校验、反调试、密钥分散存储等措施对认证密钥实施纵深防护。软件层防护方案应通过具备资质的第三方安全评测。

2） 软件层防护不应作为 L3 高风险金融与政务场景的唯一密钥保护手段，应至少与以下一项机制叠加：服务端双因子核验、动态密钥下发、硬件令牌、设备绑定证明。

c） 密钥生命周期

1） 密钥应使用符合 GB/T 32915 要求的随机数源生成。

2） 密钥更新、撤销、销毁过程应留有审计日志，L3 等级日志留存不少于 5 年。

10 认证流程规范

10.1 分级认证发起流程

业务系统判定业务风险等级，确定 L1/L2/L3 认证模式。

系统下发对应等级认证指令，明确模态、交互、防护要求。

按对应等级规范完成数据采集、核验、判定。

10.2 图视音多模态身份认证流程

身份发起：用户提交认证申请，系统匹配安全等级下发认证指令。

多模态采集：按等级要求实时采集对应模态数据，同步完成质量检测。

单模态检测：分别对各模态数据进行伪造检测、活体检测、特征提取。

跨模态核验：L2/L3 等级校验多模态数据一致性、主体同一性。

结果判定：综合核验结果，判定认证通过/失败/存疑。

后续处理：认证通过则授权访问；认证失败则拒绝服务；存疑则启动二次认证或人工复核。

11 测试与评估要求

11.1 功能测试

验证 L1-L3 全等级认证功能完整性，确保各等级模态、防护、交互要求落地有效。

11.2 安全性能评估

a） 抗伪造攻击测试：模拟 AI 图像合成、Deepfake 换脸、AI 语音合成等攻击场景，验证各等级 FPR、TPR 指标达标情况。

b） 数据安全测试：核查身份数据加密、脱敏、存储、传输安全性，杜绝数据泄露风险。

c） 系统鲁棒性测试：验证系统在网络攻击、异常数据、高并发场景下的稳定性。

d） 采集真实性测试：模拟虚拟摄像头注入、Hook 注入、跨进程帧投递、预录制视频重放等攻击场景，验证采集环节真实性识别能力。

e） 密钥保护强度测试：核查认证密钥的生成、存储、使用、销毁全生命周期安全性，验证硬件密钥库使用情况与软件层防护强度。

11.3 性能指标要求

11.3.1 响应时延要求

系统应在保障安全检测逻辑的前提下，满足以下全链路响应时间（从采集终端发起请求至返回认证结果）：

L1 级（低风险）： 验证响应时间 ≤2000 ms。

L2 级（中风险）： 验证响应时间 ≤3000 ms。

L3 级（高风险）： 验证响应时间 ≤5000 ms（含多模态一致性校验及后台专家系统/二次核验时间）。

11.3.2 抗伪造检测与比对指标执行

系统必须严格执行表 1 规定的 APCER、BPCER、FAR、FRR 及 IAPCER 等各项定量指标。在评估过程中，各指标应按 附录 B（规范性附录） 规定的标准化测试流程进行，严禁私自放宽指标阈值。

12 测试数据集

12.1 数据集构成原则

测试数据集应覆盖多样化的生物特征呈现方式、数字注入手段及真实场景环境。数据集构建应遵循公平性原则，确保不同年龄、性别、种族及地理分布的样本占比均衡，避免算法偏倚。

12.2 样本规模与统计要求

为保证性能指标在统计学上的有效性，数据集应满足以下规模要求：

a） 真实样本 (Bona Fide）： 真人尝试次数不少于 150 次（建议由至少 5-6 名受试者参与，每名受试者覆盖不同光照与动作姿态）。

b） 攻击样本 (PAI Species）： 至少覆盖 6 类典型的 PAI species（如：高清屏、打印件、硅胶面具、3D 模型、AIGC 合成样本、语音克隆音频等）。每类 PAI species 至少进行 50 次独立的攻击尝试，全集攻击样本总量不得少于 300 次。

c） 统计显著性： 上述规模设计旨在满足 95%置信区间下，有效区分 BPCER（如：5%与 10%之间的差异）。测试报告中所有性能指标（APCER, BPCER, IAPCER 等）均应同步报告其 Wilson 95%置信区间下界，以体现评估的严谨性。

12.3 数据集分类细则

a） 真实样本库 (Bona Fide Dataset）： 包含不同年龄、种族、光照条件、遮挡（如眼镜、口罩）及动态行为的真实用户样本。

b) 呈现攻击数据集 (Presentation Attack Dataset）： 包括但不限于高清显示屏重放、高分辨率照片打印、定制化 3D 硅胶面具、物理模型等。

c） 数字注入攻击数据集 (Digital Injection Dataset）： 包括主流 Deepfake 模型生成的换脸流（如 FaceSwap, SimSwap 等）、AIGC 超写实数字人视频、语音克隆音频（TTS/VC）等。

d） 篡改与对抗数据集 (Tampered & Adversarial Dataset）： 包含经深度学习修复的证照、抠图拼接图像，以及针对性加入扰动的对抗样本。

e） 极端环境数据集 (Adverse Condition Dataset）： 模拟低光照、强背景噪声、高压缩损耗、弱网络波动及多路径反射等不利环境。

12.4 数据集管理与伦理合规

a） 合规性： 真实样本的采集必须严格遵守《中华人民共和国个人信息保护法》及相关数据安全规定，所有受试者均需签署知情同意书，明确数据用途与隐私保护条款。

b） 权利保障： 伪造数据集的制作不得包含未经授权的真实人物肖像或生物特征，严禁通过非法手段获取或使用个人生物特征信息用于算法训练与评估。

c） 匿名化处理： 所有用于公开测试或第三方评测的样本，必须进行严格的脱敏或匿名化处理，防止受试者身份泄露。

12.5 负面样本比例约束

测试数据集中的攻击样本（含呈现攻击与注入攻击）总量占比不应低于总样本数的 30%，以确保评估模型对复杂攻击模式的敏感性与鲁棒性。

13 管理与责任要求

本标准涉及的身份认证**系统开发方、服务提供方、业务运营方及监管方**，应分别在系统设计、服务部署、日常运营与监督审计等环节，遵循本文件相应条款的要求。

身份认证系统运营方需建立分级安全管理制度，按业务场景适配对应认证等级。定期开展安全自查与漏洞检测，确保各等级安全指标持续达标。及时更新认证算法、安全策略，应对新型智能身份伪造技术。发生身份伪造安全事件时，立即启动应急处置，阻断风险、留存证据、上报监管部门。

身份认证系统主体应履行以下法律备案义务：

a） 算法备案义务：使用生成式 AI 辅助身份认证流程的，按《生成式人工智能服务管理暂行办法》§17 完成算法备案；

b） 网信备案义务：存储 ≥10 万人人脸信息的，按《人脸识别技术应用安全管理办法》§15 在 30 工作日内向省级网信部门备案；

c） 关键信息基础设施场景应符合《关键信息基础设施安全保护条例》§19-22 的额外要求。

违反本标准导致身份伪造安全事故、数据泄露的，依法追究相关主体责任。

14 高风险金融与政务场景专项认证细则

14.1 适用场景范围

本专项细则对应 L3 高风险等级，为强制性要求，适用于金融、政务领域所有高敏感、高风险身份核验业务，具体包括：

金融场景：银行大额转账、信贷审批、资金托管、跨境支付、数字人民币交易、证券期货开户、资产变更、密码重置、对公账户开立、反洗钱身份核验、信用卡大额消费授权等。

政务场景：社保/医保/公积金敏感信息查询与变更、户籍迁移、婚姻登记、房产过户、出入境业务办理、电子证照签发、政务电子签章、法人授权、司法核验、跨省通办敏感业务等。

14.2 基础强化认证要求

全面执行 L3 高风险等级的核验要求，禁止降低防护标准。

原则上需执行：实时视频活体检测+实时随机语音口令核验+实体证件防伪核验。

叠加环境与设备强校验：同步核验设备指纹、网络环境、地理位置、操作行为、用户风险画像，异地、新设备、夜间等高风险状态需追加核验。

全流程区块链存证，认证数据不可篡改，支持监管审计、司法举证。采用区块链技术存证时，应对原始生物特征、声纹等个人敏感信息进行去标识化处理（如仅存储哈希值或加密后的特征模板），确保链上存证内容不包含可直接识别个人身份的敏感原始数据，实现可验证性与隐私保护的平衡。

14.3 金融高风险场景专项要求

14.3.1 资金交易类业务

设定大额交易阈值，达到阈值后强制启用三模态认证+物理安全介质双重认证，物理介质包括 UKey、安全盾、数字证书、动态口令牌等。

夜间交易、异地交易、新设备首次交易、高频大额交易，自动触发二次身份重认证，必要时启动人工复核。

声纹、人脸生物特征必须与银行预留身份信息、证件信息 1:1 精准匹配，不允许模糊匹配通过。

转账、支付等核心交易环节，需增加交易信息语音/文字确认，核验用户指令真实性。

14.3.2 信贷与风控业务

借款人、担保人、共借人需独立完成完整多模态认证，不得共用、替代认证结果。

认证流程强制要求：身份证芯片/二维码防伪核验+4 秒以上连续动态活体视频+随机唇语口令核验，杜绝远程代操作、预录制视频。

信贷审批、放款环节，需联动反欺诈系统，核验身份信息与征信、风控数据一致性，发现异常立即终止认证。

14.3.3 反洗钱与账户实名制

对公账户开户、变更、网银授权、资金划转，需完成法人+经办人双主体多模态认证，同时核验企业证照、授权文书防伪信息。

账户出现异地高频登录、大额集中转入转出、可疑交易，立即触发身份重认证，连续异常则冻结账户权限。

客户身份资料需与多模态认证数据绑定留存，留存期限符合金融监管相关规定。

14.4 政务高风险场景专项要求

14.4.1 政务服务核心业务

涉及户籍、房产、社保、医保、公积金等个人敏感信息查询、变更、支取，必须完成实时活体人脸认证+身份证防伪核验+政务可信身份平台验签。涉及公积金提取的必须通过银行二次的活体认证+身份核验等多因子认证。

法定事项办理（婚姻登记、司法授权、房产过户等），多模态认证记录需上链存证，确保不可篡改、可司法采信。

政务自助终端、线上 APP 办理业务，需增加防远程操控、防翻拍检测，核验办理场景真实性。

14.4.2 电子证照与电子签章

电子身份证、电子营业执照、电子社保卡等证照签发、调用，需先完成多模态身份认证，认证通过后方可生成、调用证照。对每天的调用次数设置阈值，超过阈值的要有阻断机制，防止盗用。

电子签章、法人电子签章，实行自然人身份绑定认证，签章操作全程留痕、实时审计。

电子证照跨部门、跨区域调用，需通过政务可信身份网关完成身份互认，无需重复采集生物特征。

14.4.3 跨域政务办理

跨省通办、跨部门联办业务，身份认证结果需在可信政务网络内共享互认，保障认证安全性与便捷性。

远程政务办理业务，需核验办理人现场环境、人脸姿态，杜绝代办、伪造身份办理，建议采用多因子核验。

14.5 多模态认证强化细则

14.5.1 视频活体强化

视频采集时长不低于 4 秒，指令为系统随机动作（如摇头、眨眼、张嘴、点头）。

强制检测深度伪造、3D 面具、屏幕翻拍、预录制视频、虚拟形象等攻击手段，检测到风险直接判定认证失败。

核验视频时序连续性，杜绝剪辑、拼接、换脸等篡改视频。

14.5.2 语音认证强化

采用系统随机生成数字串、无规律短语作为朗读口令，禁止使用固定话术、自定义语音。

同步校验声纹特征、语音频谱、唇形与语音同步性，精准识别 AI 合成语音、声音克隆攻击。

语音采集过程实时监测环境噪声、静音、循环播放等异常状态，异常则终止采集。

14.5.3 证件核验强化

支持身份证、营业执照等证件全维度防伪检测，包括缩微文字、荧光纹理、防伪膜、印章锯齿、底纹一致性等。

精准识别 PS、拼接、涂改、伪造印章等证件篡改痕迹，过期、挂失证件直接判定无效。

支持证件芯片读取、二维码验真，实现线上线下身份信息一致性核验。

政务及金融实名场景应优先对接国家网络身份认证公共服务（CTID / IDaaS）完成实名验真，减少重复采集与本地留存。

14.5.4 跨模态一致性强制校验

出现以下任一情形，直接判定认证失败，触发风险告警：

a） 人脸与证件人像匹配度不达标

b） 语音内容与唇形动作完全不一致

c） 视频动作与系统指令不匹配

d） 设备、地理位置与常用环境异常

e） 单模态检测出明显伪造特征

14.6 安全审计与应急处置

14.6.1 日志与存证管理

L3 高风险场景认证日志、操作记录、多模态核验数据至少留存 5 年，日志内容包含时间、地点、设备、网络、操作内容、各模态核验结果、算法版本、风险判定结果。

核心业务认证记录、电子签章、资金交易认证记录，必须采用区块链存证，支持第三方机构审计、司法出证。

14.6.2 风险应急处置

连续认证失败≥3 次、检测到 Deepfake/AI 合成攻击、跨模态核验异常，立即冻结用户操作权限，推送实时风险告警。

发生重大身份伪造攻击事件，1 小时内上报对应金融/政务监管部门，24 小时内提交应急处置报告，留存全部攻击证据。

建立可疑认证记录回溯机制，支持离线二次检测、第三方安全机构复核鉴定。

附 录 A

（资料性附录）

智能身份伪造攻击案例分析

本附录提供了典型的智能身份伪造攻击案例，用于帮助理解和实施本文件规定的安全要求。

案例一

2024 年 1 月，国际知名工程顾问公司奥雅纳（Arup）香港分公司遭遇了一起极具代表性的 AI 深度伪造（Deepfake）诈骗案。诈骗分子预先通过公开渠道搜集了公司高层管理人员的影像与音频资料，利用先进的 Deepfake 技术合成了 CFO 及其他同事的逼真数字分身。在精心策划的“多人视频会议”中，仅有一名财务员工是真人，其余参会者均为 AI 生成的虚拟形象，从而构建了一个极具迷惑性的虚假权威场景。

受此诱导，该员工在缺乏有效核实的情况下，分 15 次向指定账户转账共计 2 亿港元（约合 2500 万美元）。此案深刻暴露了传统身份验证手段在生成式 AI 面前的脆弱性，标志着网络犯罪已正式进入多模态伪造的新阶段。它强烈警示企业必须针对高价值交易建立严格的“带外验证”及多重审批机制，以防范日益逼真的社会工程学攻击。

案例二

2024 年 4 月，福建福州某科技公司法人代表郭先生在 10 分钟内遭遇了一起极具迷惑性的 AI 换脸与声音克隆诈骗案，涉案金额高达 430 万元。诈骗分子首先通过技术手段盗取了郭先生好友的社交账号，并利用深度合成技术（Deepfake）精准克隆了该好友的面部特征与声音。

在微信视频通话中，骗子伪装成郭先生的好友，以“朋友外地竞标急需保证金，需借用公司账户走账”为由，出示了伪造的转账底单截图。基于对熟人面孔、声音的信任以及对“公对公走账”话术的放松警惕，郭先生未核实资金实际到账情况便匆忙转账。所幸经两地警银联动紧急止付，成功拦截了 336.84 万元，但仍有 93.16 万元资金被转移。

此案是典型的利用 AI 多模态技术实施的精准社会工程学攻击。它深刻警示公众，在“眼见不一定为实”的数字时代，传统的视频核实已不足以完全确认真实身份。面对涉及资金的敏感请求，必须通过电话回拨、要求对方做特定动作（如摸鼻子、眨眼）等多重渠道进行交叉验证，严防生物信息泄露与新型技术诈骗。

案例三

2024 年 4 月，北京互联网法院审结并宣判了全国首例 AI 生成声音人格权侵权案。原告殷某作为一名专业配音演员，发现其声音被某文化传媒公司违规授权给某软件公司，后者利用人工智能技术对其声音素材进行深度学习与合成，开发出文本转语音（TTS）产品并对外出售。该产品在相关平台上的播放量高达 32.6 亿次，严重侵害了原告的声音权益。

法院审理认为，自然人声音具有独特性和人身专属性，受法律保护。经当庭勘验，涉案 AI 声音在音色、语调及发音风格上与原告高度一致，具备明确的“可识别性”，因此自然人的声音权益保护范围应及于 AI 生成声音。最终，法院依法判决侵权方赔礼道歉并赔偿经济损失 25 万元。

此案作为最高法发布的典型案例，具有里程碑式的司法意义。它不仅明确了 AI 合成声音的法律属性，更厘清了录音制品著作权与声音人格权的界限，为人工智能时代的声音数据合规与技术向善划定了清晰的法律红线。

案例四

2024 年 8 月，一起极具代表性的“AI 马斯克”投资骗局被媒体曝光，引发了全球对深度伪造（Deepfake）技术滥用的高度关注。该案中，诈骗分子截取特斯拉 CEO 埃隆·马斯克的真实采访视频，利用先进的唇形同步技术与语音克隆工具，伪造出带有其标志性南非口音的虚假投资广告，并承诺“快速且无风险的丰厚回报”。

一名 82 岁的美国退休老人史蒂夫·比彻姆（Steve Beauchamp）在观看该视频后深信不疑，联系了广告背后的虚假外汇公司。在骗子的诱导下，他先期投入 248 美元试水，随后在数周内陆续耗尽毕生积蓄，累计转账高达 69 万美元（约合人民币 495 万元），最终血本无归。

此案深刻揭示了深度伪造技术“公信力嫁接”的犯罪特征——即利用名人的社会影响力大幅降低受害者的心理防线。同时，这也暴露出社交媒体平台在虚假广告审核与 AI 内容监管方面的重大漏洞。该案件警示公众，在数字时代面对看似权威的视频背书时，必须保持审慎，通过官方渠道交叉核实信息，严防陷入 AI 技术编织的金融陷阱。

案例五

2025 年 9 月，iProov 威胁情报团队披露了一款针对越狱 iOS 设备的新型复杂攻击工具，该工具最早于 2024 年 12 月在野外被发现。这一发现标志着数字身份欺诈手段在工业化与程序化方面取得了重大突破。

该工具专门针对 iOS 15 及更高版本的系统，其核心攻击逻辑在于利用越狱设备被破坏的安全架构。

攻击者通过远程呈现传输机制（RPTM）服务器，在本地计算机与被攻陷的 iPhone 之间搭建起数据桥梁，从而将精心制作的深度伪造内容（如换脸视频或动作重演）直接注入设备的视频流中。这种“数字注入”手段完全绕过了物理摄像头硬件，能够欺骗应用程序，使其误将合成媒体识别为真实的实时生物特征。

该案例深刻揭示了传统身份验证方法在面对底层系统篡改时的脆弱性。它强烈警示安全行业，单纯的活体检测已不足以应对威胁，必须部署包含官方文件匹配、恶意媒体元数据分析以及被动挑战－响应交互在内的多层防御体系，以应对日益严峻的合成身份欺诈风险。

附 录 B

（规范性附录）

测试方法

本附录规定了本标准涉及的各项安全要求的验证方法。所有安全测试应由具备相应环境模拟能力的第三方安全实验室开展，确保测试数据的独立性与客观性。测试记录应完整留存，保存期限不少于 5 年。

B.1 业务逻辑攻击模拟测试

旨在评估系统在处理复杂业务流程中的安全性，测试应包括：

a） 凭证重放攻击测试：通过捕获历史认证数据包并重放，验证系统抗重放能力；

b） 中间人攻击测试：模拟网络层中间人，尝试劫持、篡改认证报文，验证加密传输与签名有效性；

c） 提示注入诱导测试：在交互环节输入恶意指令，诱导认证系统执行非预期操作，验证系统输入合法性校验能力；

d） 身份冒用攻击测试：尝试使用合法但非本人持有的认证凭证（如数字证书、硬件令牌）进行身份冒用，验证多因子绑定机制。

B.2 认证强度与密钥安全评估测试

旨在验证身份核验系统的核心密码学防护能力，应包括：

a） 密码学算法合规性测试：验证签名、加密、哈希算法是否符合 8.1.1(c)(2)及国家密码管理局相关要求。

b） 密钥管理安全性测试，包括：

1）硬件密钥库使用核查：通过设备完整性证明接口（如 Android Key Attestation、iOS App Attest 或国产可信执行环境接口）核验密钥是否在硬件密钥库内生成、存储、使用。L3 等级密钥应 100% 在硬件密钥库内完成全生命周期操作；

2）软件层防护评估：对不具备硬件信任根的实施环境，应通过静态逆向分析、动态调试、内存提取等方式评估密钥防护强度。评估应由具备 CMA 或 CNAS 资质的第三方安全机构执行，密钥提取攻击成本应满足相应安全等级要求；

3）密钥生命周期审计：核查密钥生成、更新、撤销、销毁全过程的审计日志完整性。L3 等级日志

留存期应不少于 5 年，且应能溯源到具体操作时间、操作主体、操作结果。

c) 审计日志可追溯性测试：验证认证全流程日志的完整性、防篡改能力及检索效率。

B.3 安全采集验证测试

a） 设备完整性证明有效性测试：核验采集终端提交的完整性证明可由认证服务方独立验证；模拟 Root/越狱、应用重打包、调试器附加、模拟器运行等场景，验证证明机制能识别并拒绝异常环境。L3 级应强制通过硬件信任根证明测试。

b） 受控采集管线防注入测试：部署主流虚拟摄像头、虚拟麦克风驱动，验证系统能否识别并终止采集；通过 Hook 技术替换系统原生采集接口，验证系统是否具备检测能力并拒绝非原生数据流；模拟恶意应用向采集进程注入伪造视频帧，验证系统能否识别帧来源异常；篡改陀螺仪、光照等辅助信号使其与视频内容物理矛盾，验证系统交叉校验能力。L3 级应通过全部四项。

c） 采集时签名与时间戳验证测试：在采集会话完成后篡改采集数据、传感器辅助信号或时间戳字段，验证签名验证机制能 100% 检出篡改。核验时间戳来源于具备国家授时资质的时间戳服务机构（TSA）；模拟设备本地时间篡改、TSA 响应伪造等场景，验证时间戳机制不被绕过。在采集端以外的独立环境验证签名有效性、证书链可信性，验证应不依赖采集端任何运行时支持。各项测试样本数不少于 30 次，应 100% 检出篡改与异常环境。

B.4 呈现攻击检测（PAD）测试方法

B.4.1 测试架构

进行 PAD 测试时，应明确测试范围： PAD 子系统级测试仅测试呈现攻击检测模块；数据捕获子系统级测试应测试从采集到检测的完整链路；整体系统级测试应端到端测试完整认证流程。

B.4.2 PAI species 分级

PAI species 应按制作难度和成本分为三档：

Level 1（普通用户级）：制作成本 ≤30 美元，如打印照片、手机重放

Level 2（进阶级）：制作成本 ≤300 美元，如高清屏幕、2D 面具

Level 3（红队级）：无成本限制，如定制 3D 面具、AIGC 合成样本

测试结果应与 iBeta Level 1/2/3、FIDO Level A/B 兼容。

B.4.3 样本规模要求

每个 PAI species 测试次数不少于 50 次，真人尝试次数不少于 150 次。

B.4.4 指标报告格式

APCER/BPCER 报告格式应按 ISO/IEC 30107-3:2023 §10 要求执行：按 PAI species 分别报告错误率取最差 PAI species 的指标作为最终结果，在固定 APCER 工作点（如 10%、5%）报告对应的 BPCER 值

B.5 注入攻击与合成内容检测测试方法

聚焦于对非物理介质产生的数字伪造流检测，

B.5.1 测试架构

本项测试主要针对系统对数字注入流的识别能力，测试对象应覆盖：a) 合成身份内容（Synthesized Identity Content）： 由 AIGC 模型直接生成的深度伪造人脸视频、数字人形象、AI 克隆语音等。b) 远程数字注入攻击（Remote Digital Injection）： 利用 RPTM（远程呈现传输机制）或跨进程流投递方式，绕过真实摄像头物理采集，直接将处理好的数字流馈送至系统前端。

B.5.2 注入向量分类（IAPCER 测试项）

测试应覆盖以下五类典型的数字内容注入向量，并按类别分别评估 IAPCER 指标：

IA-1 合成人脸注入： 测试系统对基于 GAN/扩散模型生成的超写实人脸视频流的识别能力。

IA-2 语音克隆与合成注入： 测试系统对使用 TTS（文本转语音）或 VC（语音转换）技术生成的伪造音频流的识别能力。

IA-3 远程流转发（RPTM）： 测试系统对通过网络协议（如 RTMP/WebRTC）远程传输的虚假媒体流的检测能力。

IA-4 视频序列篡改与拼接： 测试系统对人脸换脸（Deepfake）、视频背景重构及面部关键点重合等数字篡改行为的检出能力。

IA-5 语义与时序异常注入： 测试系统识别音频内容与唇形轨迹不匹配、说话韵律异常等逻辑性伪造的能力。

B.5.3 测试实施流程

环境隔离： 测试应在“受控采集管线”（B.3.2）通过验证的基础上进行，以确保测试结果反映的是“检测算法对数字内容特征的识别能力”，而非采集管线的拦截能力。

攻击触发： 对认证系统发起上述五类注入攻击，攻击载荷应包含多种复杂环境下的变体（如不同压缩比、不同帧率、不同噪声等级）。

判定指标： 记录系统对每种注入向量的 IAPCER，并根据 L1-L3 的等级要求，对比是否满足对应阈值。

B.5.4 性能要求

测试应涵盖经过不同编码格式（如 H.264, H.265, AAC 等）压缩后的攻击样本， 每类向量测试次数不少于 30 次，总测试次数不少于 150 次。

B.6 多模态联合测试方法

B.6.1 跨模态一致性测试

异常样本应覆盖以下 4 类： 1）人脸真实但证件伪造 2）证件真实但人脸伪造 3）视频真实但音频

伪造 4）音视频均为伪造但来源不同主体

每类异常样本不少于 50 例，指标报告包括跨模态异常检出率。

B.6.2 唇语－语音同步检测

采用 SyncNet 类模型进行检测，当音视频偏移 > 100 ms 时检出率应 ≥ 98%。

B.7 测试报告格式要求

测试报告应包含以下必填字段：

a） 系统标识（名称、版本、厂商）

b） 测试范围（PAD/数据捕获/整体系统）

c） 设备清单（测试终端型号、操作系统版本）

d） PAI/IA 清单（使用的攻击工具种类及规格）

e） 各类错误率（按 PAI species 和注入向量分别报告 APCER/BPCER/IAPCER）

f） Wilson 95% CI 下界（每项指标的置信区间）

g） 失败示例可视化与归因（样本图片及判定错误原因）

h） 有效期（报告有效期 ≤ 12 个月）

参 考 文 献

[1] GB/T 25069-2022 信息安全技术 术语

[2] GB 45438-2025 网络安全技术 人工智能生成合成内容标识方法

[3] GB/T 22239-2019 信息安全技术 网络安全等级保护基本要求

[4] GB/T 35273-2020 信息安全技术 个人信息安全规范

[5] GB/T 41819-2022 信息安全技术 人脸识别数据安全要求

[6] GB/T 38671-2020 信息安全技术 远程人脸识别系统技术要求

[7] GB/T 41807-2022 信息安全技术 声纹识别数据安全要求

[8] GB/T 42981-2023 信息技术 生物特征识别 人脸识别系统测试方法

[9] GB/T 44248-2024 信息技术 生物特征识别 人脸识别系统应用要求

[10] GB/T 41772-2022 信息技术 生物特征识别 人脸识别系统技术要求

[11] GB/T 37078-2022 信息安全技术 生物特征识别安全技术要求