ICS 35.040
L 71

团　　体　　标　　准

T/AI 110.1—2020

人工智能视觉隐私保护 第1部分：通用技术要求

Visual privacy protection of artificial intelligence——

Part 1: General technical requirements

2020－12－31 发布　　2020－12－31 实施

中关村视听产业技术创新联盟　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。

T/AI 110 在《人工智能视觉隐私保护》的总标题下，包括以下 3 个部分：

——第 1 部分：通用技术要求；

——第 2 部分：技术应用指南；

——第 3 部分：评估方法。

本文件为 T/AI 110 的第 1 部分。

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专利的责任。

本文件由新一代人工智能产业技术创新战略联盟提出。

本文件起草单位：青岛海信电子产业控股股份有限公司、青岛海信网络科技股份有限公司、天津大学、山东大学、云从科技集团股份有限公司、科大讯飞股份有限公司、中国电子技术标准化研究院。

本文件主要起草人：陈维强、高雪松、李克秋、李玉军、胡伟凤、张淯易、孙菁、孟祥奇、孙宗臣、温浩、李军、马万钟、吴子扬、余明明、李婧欣、曹策。

引　言

人工智能技术已对人类社会生活的各个方面产生了深远的影响，如专家系统、智能机器人、机器翻译等在交通、医疗等领域给人类带来极大的便捷。但如同任何一项新技术一样，人工智能技术在给人类带来诸多便利的同时，也引发了一些问题，比如个人隐私保护问题。

个人隐私在实践中也被称为个人隐私数据或个人信息。长久以来，个人隐私保护都是各国立法关注的核心议题。随着人工智能迅速发展，个人隐私保护问题俨然成为各方密切关注的重点，特别是伴随着大数据技术和人工智能技术的结合，政府和企业的决策不断加深对大数据分析的依赖。大规模的数据收集、分析和使用，使传统社会走向透明化，在万物互联、大数据和深度学习三者叠加后，个人隐私将会成为“奢侈品”，人类将不再有隐私可言。

在如今的人工智能时代，数据从采集到使用等各个环节都面临着新的风险。在数据采集环节，大规模设备自动收集着成千上万的用户数据，其中涉及到的用户人脸、姿态等视觉数据在海量汇聚后可形成对用户的全面跟踪。在数据使用环节，人工智能技术通过用户浏览及搜索记录、设备信息、位置信息、订单信息，提取用户的浏览、搜索偏好、行为习惯、位置信息等特征，分析出深层信息，进一步扩大了用户隐私暴露的风险。在整个数据的生存周期中，由于系统安全漏洞等原因，个人视觉隐私数据还面临着被泄露的潜在安全风险。

技术更新迭代是推动社会进步的重要力量，不能因为隐私问题阻碍人工智能技术继续前进，但人工智能的发展也不能以牺牲隐私权为代价。因此亟需提出方案对人工智能视觉隐私数据进行保护。

人工智能视觉隐私保护 第1部分：通用技术要求

1 范围

本文件规定了人工智能技术中涉及视觉隐私保护的通用技术要求，给出了隐私数据控制者的义务、保障措施等约束行为和视觉隐私数据保护定级指南。

本文件适用于规范涉及人工智能视觉隐私数据的应用、服务、系统应当进行的隐私保护处理，也适用于指导本领域技术人员在人工智能视觉相关的开发中开展隐私保护方案设计和实施。

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（包括所有的修改单）适用于本文件。

GB/T 25070 信息安全技术 网络安全等级保护安全设计技术要求

GB/T 35273 信息安全技术 个人信息安全规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

人工智能视觉隐私数据 AI visual privacy data

简称视觉隐私数据，人工智能视觉隐私数据主体的隐私信息，如用于人脸识别、姿态识别等相关的感知和行为数据。

3.2

处理 process

对视觉隐私数据的一个或一系列操作，诸如收集、记录、存储、检索、使用、披露、传输等。包括但不限于将视觉隐私数据上传至服务器进行处理的行为和在端侧、边缘侧等非云端设备中对视觉隐私数据进行处理的行为。

3.3

控制者 controller

处理视觉隐私数据的自然人、法人、公共机构、行政机关、企业等。

3.4

第三方 the third party

就所涉及的问题而言，公认与相关各方均独立的个人或团体。本文将特指视觉隐私数据主体、控制者以及在控制者授权处理视觉隐私数据者以外的自然人、法人、行政机关等。

3.5

同意 consent

视觉隐私数据主体的“同意”，指视觉隐私数据主体自愿做出的，允许视觉隐私数据被控制者处理。

3.6

视觉隐私数据泄露 visual privacy data disclosure

违反信息安全策略，使视觉隐私数据被未经授权的实体使用。

3.7

代表 delegate

代表控制者履行本文件规定义务的自然人或法人。

3.8

生物特征数据 biometric data

用户生物学的和行为的特征，该特征可被检测，并且可以从中提取有区别的、可重复的生物特征项，从而达到个体自动识别的目的。

3.9

安全芯片 secure chip

含有密码算法、安全功能，可实现密钥管理机制的集成电路芯片。

3.10

采集安全 collection secure

一种将各类视觉传感器等终端采集设进行安全采集、存储和转发的方法。

3.11

密箱 secure box

一种具备不可拆卸、不可篡改等安全性高的独立存储密钥的软硬件装置

3.12

双因子 double factor

两个影响某事件的因素

3.13

多因子 multi-factor

多个影响某事件的因素

3.14

第三方软件 third party software

第三方提供的服务，系统或其他形式的软件装置。

3.15

应用层 application layer

提供服务的应用，应用层通常通过对感知层采集的数据进行特征提取、数据分析、计算等，形成用于人工智能服务的结果。

3.16

传输层 transport layer

系统中负责数据流传输等用于提供通信功能的服务。

3.17

感知层 perceptual layer

使用传感器采集音视频和其他类型数据的服务和相关系统。

4 人工智能视觉隐私数据控制者

4.1 义务

视觉隐私数据控制者应确保并能够证明，对视觉隐私数据的处理是按照本文件进行的，并保证数据主体实现相应的权利。

视觉隐私数据控制者应为视觉隐私数据主体提供方法来访问和控制用户的视觉隐私数据。

视觉隐私数据控制者应确保视觉隐私数据处理安全，包括但不限于：

a) 在处理视觉隐私数据的过程中，视觉隐私数据控制者必须采取相应的技术措施（如加密特征编码）来保证视觉隐私数据的安全；

b) 视觉隐私数据控制者和处理者在评估安全加密算法级别时，必须考虑处理过程中所带来的可能风险，比如非法破坏，遗失，未经授权披露或访问传输；

c) 视觉隐私数据控制者必须采取措施，确保视觉隐私数据不被其他人处理，除非控制者指示或者法律要求。

4.2 安全保障措施

人工智能视觉隐私数据控制者应采取技术措施保障视觉隐私数据的安全，包括但不限于：

a) 对视觉隐私数据的匿名化；

b) 对视觉隐私数据的存储和处理确保提供针对性的保护方式；

c) 建立定期测试、评估、评价技术和管理措施是否有效的体系。

4.3 视觉隐私数据的泄露报告

如果视觉隐私数据主体的视觉隐私数据被泄露，控制者应在72小时以内向视觉隐私数据主体通知并提供相关信息，除非该泄漏不会对视觉隐私数据主体的权益造成侵害。

针对泄露报告，控制者至少应提供以下信息：

a) 被泄露的视觉隐私数据类型；

b) 造成视觉隐私数据泄露的原因；

c) 控制者采取的措施。

5 人工智能视觉数据隐私保护框架

人工智能视觉数据隐私保护框架分为风险评估、安全技术、安全管理三个维度，如图1所示。在安全技术层面，具有处理人工智能视觉数据行为的产品应在采集安全、传输安全、存储安全、处理安全、应用安全5个方面部署相应安全技术手段；在安全管理层面，供应商应在用户管理、权限管理、安全审计、容灾备份等方面部署相应技术手段；在风险评估层面，供应商必须对具有处理人工智能视觉数据行为的产品应进行安全评估、隐私评估和等级评测。

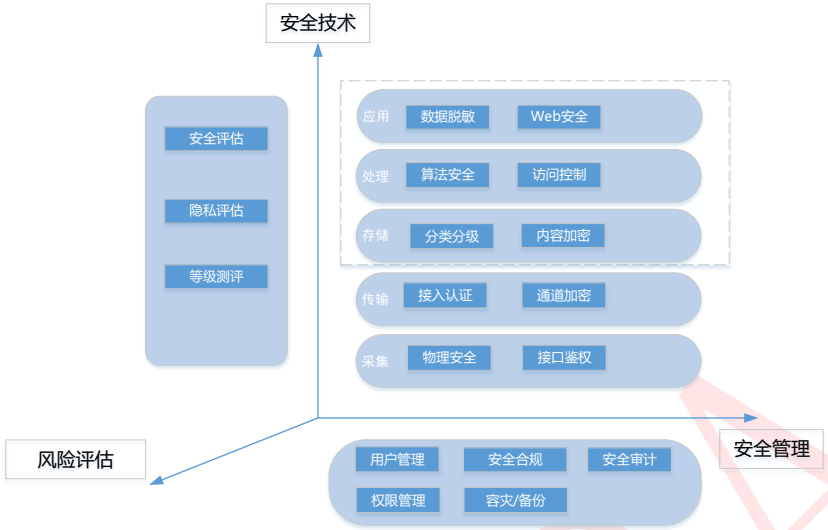

图1 视觉数据隐私保护框架

6 视觉隐私数据保护的技术要求

6.1 视觉数据隐私保护基本要求

本文件所采取的视觉数据处理应按照合法、正当、必要的原则，具体包括权责一致、目的明确、选择同意、最小必要、公开透明、确保安全、主体参与，参照GB/T 35273。

本文件中涉及的加密算法应符合国家标准设计要求，所采取的视觉数据处理应在满足“GB/T 25070 通用设计要求”中第三级要求的基础上，满足本文件的相关要求。

6.2 通用技术要求

6.2.1 采集安全要求

采集安全应确保数据采集的服务安全和终端设备的物理安全，包括但不限于：

a) 具备安全芯片、可信赖的计算环境等保障生物认证类视觉数据安全性的方案；
b) 减少设备向系统外部传递信息，关闭无用端口，包括但不限于：前期的调试接口、程序烧录及诊断测试接口；
c) 采集设备应具有物理安全防护措施，包括但不限于防干扰、防拆卸等措施，确保数据采集终端设备的物理安全。
d) 对于业务必须开放的接口，采取接口鉴权机制，防止接口被非法调用；
e) 定期审查配置设备，设备具备固件自动升级能力，且采取安全的更新方式；
f) 具备视觉数据分类和敏感隐私数据检测能力；
g) 具备物理遮蔽手段，能够通过镜头盖、旋转、伸缩等物理方式在关闭摄像服务时提供遮蔽。

6.2.2 传输安全要求

感知设备和应用之间的网络与传输安全，应满足：

a) 接入认证
   1) 应当利用安全插件进行终端异常分析等，实现终端入侵防护，避免发生借助终端攻击网络关键节点等行为；

2） 强制进行单向/双向认证机制，阻止非法节点接入；

b） 通道加密，视觉数据的传输安全原则应当满足；

1） 支持但不限于采用专用内网、安全协议的方式保证数据的安全传输方式；

2） 实现加载内容的过滤和访问限制；

3） 采用视觉加密算法传输加密后的视频数据，保证数据机密性；

4） 具有数据的完整性及时效性检验功能。

6.2.3 存储安全要求

系统在收集和使用视觉数据时，首先对数据进行分类分级，然后使用视觉加密算法加密存储视觉隐私数据。应满足如下要求：

a） 视觉数据的存储原则应当满足 6.4 中视觉数据隐私保护系统等级的相关要求；

b） 用户按照存储需要，将不同的信息存储在不同的密箱中，且密箱之间是相互隔绝的，对其他用户不可见；

c） 系统中的应用卸载或账户注销后，系统应当具备删除内部存储的所有文件的能力；

d） 外部存储可用于数据共享，开发者应谨慎使用外部存储，避免将用户的隐私数据写入外部存储；

e） 使用加密机或者密钥托管服务管理用户主密钥和数据密钥；

f） 存储用户视觉隐私数据的设备应具备对隐私数据快速加密的能力。

6.2.4 处理安全要求

处理组件应提供相应的身份鉴别和访问控制机制，确保只有合法的用户或应用程序才能发起数据处理请求。

a） 算法安全，主要面临算法黑箱、算法模型缺陷等风险，应满足如下要求：

1） 对核心算法的框架和组件进行严格的测试管理和安全认证，减少因算法漏洞和后门等引发安全风险；

2） 充分评估算法潜藏偏见和歧视，避免产生与预期不符甚至伤害性结果，确保系统决策结果可控；

3） 应保证算法的准确性，鲁棒性和可解释性，保证算法的核心指标是可验证的；

4） 算法应具备一定的防对抗攻击能力，以防止经过特殊构造产生的样本让分类器的分类结果不可靠；

5） 加强 AI 模型的保护，采用高强度的加密算法对 AI 模型进行保护，或者采用其他更高级的数据保护机制进行保护。

b） 身份鉴别与访问控制，应满足如下要求：

1） 对于 FTP 服务密码、登录密码、外部系统接口认证密码等隐私数据，应加密存储；

2） 应具备对第三方软件访问数据权限的控制能力，能够发现或记录非授权应用访问数据；

3） 对隐私级别高的视觉数据设置双因子或多因子的身份鉴别机制；

4） 对数据库账号异地登陆的情况给予提示；

5） 确保参与模型训练的数据集的内容安全，并保障数据集的完整性和代表性，防止数据投毒攻击；

6） 对生物特征数据所有用途、目的和方式都应提供可靠的安全手段，包括但不限于独立存储、数据加密、本地处理、拒绝上传等。

6.2.5 应用安全要求

人工智能视觉隐私数据的应用安全应满足如下要求：

a) 构建完善的应用管控保障机制，提供应用签名、内存保护、恶意网址检测、流量监控等措施保障应用安全；
b) 数据发布前，应当对隐私数据采取去标识化、匿名化等技术实现数据脱敏；
c) 敏感视觉数据在存储和应用时应支持脱敏；
d) 应具备代码安全审计功能，尤其防止反编译导致密钥丢失；
e) 人工智能产品或应用在保证信息安全的同时，保证功能安全；
f) 数据共享时，应采用安全共享手段，包括但不限于同态加密、联邦学习等方式；
g) 已发布或已授权共享的数据，应支持用户数据的可溯源、数据流通过程的可监控。

6.3 安全管理要求

视觉隐私数据控制者应满足如下要求：

a) 遵守隐私政策和相关规程，对安全措施的有效性进行审计；
b) 完善的用户管理机制，对用户（包括管理员）进行合理授权；
c) 应建立备份与恢复管理相关的安全管理制度，对备份信息的备份方式、备份频度、存储介质和保存期等进行规范；
d) 至少一周进行一次安全检测，包括但不限于 Web 威胁扫描、渗透测试，查找系统漏洞、研判是否挂马，及时更新安全补丁，以便快速修复漏洞；
e) 对收集的数据进行统计、分析，定期形成系统安全态势分析报告。

6.4 风险评估要求

6.4.1 视觉隐私系统定级

建立全面的视觉数据隐私保护系统，该系统分层分级的对数据进行隐私保护。按照保密程度，视觉数据隐私保护分为三级：

a) 一级视觉数据隐私保护系统
b) 二级视觉数据隐私保护系统
c) 三级视觉数据隐私保护系统

每级分别从感知层、传输层及应用层等过程，每级都必需满足6.2的通用技术要求。同时，对隐私数据的提出要求。具体参照附录B.1。

6.4.2 视觉隐私系统的隐私评估

从隐私数据、算法/模型、应用等多维评估指标、提供一套的完善的隐私评估指南。保证用户隐私数据可度量，可追踪溯源等。

6.4.3 视觉隐私系统的安全评估

包括隐私系统的产品安全、安全能力评估和安全管理等多维度进行安全评估。

产品安全评估：从隐私系统的设计、开发、测试、应用等产品开发测试过程进行安全评估。

安全能力评估：从系统的安全能力进行安全评估，如数据的保密性、算法的安全性、网络安全等方面。

安全管理评估：从管理层面进行安全评估，主要包括各项操作是否符合规范等合规性要求。

6.5 其他要求

应针对涉及未满14周岁的儿童的视觉隐私数据采取专门措施，确保：

a) 只有在征得父母或者监护人的同意等情况下才可以处理；

b) 在父母或监护人未同意的情况下收集的儿童人工智能视觉隐私数据应立即予以删除。

附 录 A
（规范性）
视觉数据隐私保护定级指南

A.1 一级视觉数据隐私保护系统

一级视觉数据隐私保护系统主要用于隐私程度最高的场景，隐私程度高于二级场景，包括但不限于采集自或内容涉及家庭内部场景的数据。该系统主要针对应用场景需求，对原始图像数据进行处理，生成结构化数据，结构化数据主要包含身份、人体姿态、人员属性、行为、事件等用户信息，系统也可针对用户要求对结构化数据进行处理，只保留用户同意留下的信息。此系统只可获取处理后的结构化数据，避免原始图像数据的泄露。

a) 应用层：用户只可获取加密后的结构化数据；
b) 传输层：对于内部传输进行加密操作，保护缓冲区的原始数据；
c) 感知层：在非用户主动行为干涉下，不得存储原始数据，只对处理后的结构化数据进行加密存储。

A.2 二级视觉数据隐私保护系统

二级视觉数据隐私保护系统主要用于隐私程度较高的场景，隐私程度低于一级场景高于三级场景，包括但不限于家庭视频通话的数据。该系统可针对不同的应用场景需求，对原始图像数据进行虚化操作或结构化操作。此系统可获取虚化后的数据以及结构化数据，并支持以去标识处理后数据的形式传送。

a) 应用层：用户可通过密钥获取加密后的虚化数据、结构化数据；
b) 传输层：对于内部传输进行加密操作，保护缓冲区的原始数据；
c) 感知层：在非用户主动行为干涉下，不得不存储原始数据，对处理后的虚化数据和结构化数据进行加密存储。

A.3 三级视觉数据隐私保护系统

三级隐私保护系统主要用于一般保密程度的场景，隐私程度低于二级场景，包括但不限于社区、城市安防场景、公开视频会议等场景中的数据。该系统支持对原始数据的加密储存，可有条件的获取原始图像数据。

a) 应用层：在查看数据的指定应用端内置解密算法，用户可通过授权后，可获取加密后的原始图像数据，使用解密算法解密后可获取原始图像数据；
b) 传输层：对于内部传输进行加密操作，保护缓冲区的原始数据；
c) 感知层：对原始图像数据进行加密处理后，可存储原始数据。