ICS

CSS

中华人民共和国国家标准

|  |
| --- |
| **GB/T** XXXXX—XXXX |

工业设计 智能功能安全设计指南

Industrial design——Guidelines for safety design of intelligent function

|  |
| --- |
| （草案稿） |
|  |

YYYY - MM - DD实施

YYYY - MM - DD发布





目次

前言 II

引言 III

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 缩略语 3

5 智能功能及安全分类 4

6 总体原则 4

7 安全设计流程 4

8 设计前准备 5

9 危害识别 6

10 安全设计 7

11 评估与确认 9

附录A 10

附录B 11

附录C 12

参考文献 15

 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专利的责任。

本文件由全国工业设计基础标准化工作组（SAC/SWG 31）提出并归口。

本文件起草单位：

本文件主要起草人：

 引言

随着智能技术的高速发展，智能产品的市场保有量逐步扩大，在给用户带来全新使用体验的同时，也对产品安全提出了新挑战。对智能产品的智能功能开展安全设计，是保障智能产品的智能功能正常运行、防范产品质量安全问题的有效手段。为了帮助智能功能安全设计工作的开展，增进产品设计人员对智能功能安全风险的理解和认识，有效消除或减小智能产品智能功能带来的安全风险，有必要建立一套通用的智能功能安全设计方法，从而为智能产品设计人员提供智能功能安全设计的总体思路和步骤。

工业设计 智能功能安全设计指南

# 范围

本文件给出了智能功能及安全的分类，安全设计总体原则，安全设计流程以及“设计前准备”、“危害识别”、“安全设计”、“测试、评估与确认”的指南。

本文件适用于指导各类组织开展产品智能功能的安全设计。

# 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，仅该日期对宜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/T 1002 家用和类似用途单相插头插座 型式、基本参数和尺寸

GB/T 2099（所有部分） 家用和类似用途插头插座

GB/T 4706（所有部分） 家用和类似用途电器的安全

GB 4943.1 音视频、信息技术和通信技术设备 第1部分：安全要求

GB/T 7000（所有部分） 灯具

GB/T 15706 机械安全 设计通则 风险评估与风险减小

GB/T 16915（所有部分） 家用和类似用途固定式电气装置的开关

GB/T 25295 电气设备安全设计导则

GB/T 27921 风险管理 风险评估技术

GB/T 32921 信息安全技术 信息技术产品供应方行为安全准则

GB/T 35273 信息安全技术 个人信息安全规范

GB/T 41295.1 功能安全应用指南 第１部分：危害辨识和需求分析

GB/T 41295.2 功能安全应用指南 第2部分：设计和实现

# 术语和定义

下列术语和定义适用于本文件。

##

智能 intelligence

具有人类或类似人类智慧特征的能力。

注：人类或类似人类的智慧特征，表现为在实现某个目的的过程中，总会经历一个或多个的感知、决策、执行的过程或过程循环，并在其中通过不断学习，提高自身实现目的的能力和实现目的的效率与效果：本文件认为，在体现人类或类似人类的智慧特征上，感知、决策、执行和在其中的学习的各项能力和过程具有不可或缺性。

[来源：GB/T 41790-2022，定义3.2]

##

智能产品 intelligent products

应用了智能技术或具有了智能功能的产品。

[来源：GB/T 41789-2022，定义3.1，有修改]

##

智能功能 intelligent function

通过应用智能技术，使产品在特定场景中表现出人类或类似人类智慧特征的能力集合，其核心是通过感知、决策、执行和学习环节，动态优化过程或结果。

##

智能技术 intelligent technologies

使产品或事物具备人类或类似人类智慧特征的技术或技术解决方案。

[来源：GB/T 28219-2018，定义3.7，有修改]

##

安全设计 safety design

在产品设计阶段中，通过系统性方法分析并预防危害发生、降低风险，并满足智能功能安全要求的设计过程。

##

伤害 harm

对人体健康的损害或损伤，对财产或环境的损害。

[来源：GB/T 20002.4-2015，定义3.1]

##

危害 hazard

可能导致伤害的潜在根源。

[来源：GB/T 20002.4-2015，定义3.2]

##

风险 risk

伤害发生可能性和伤害严重程度的组合。

[来源：GB/T 34400-2017，定义2.14]

##

可容许风险 tolerable risk

按当前社会价值取向在一定范围内可以接受的风险。

注：在本标准中，“可接受风险”和“可容许风险”为同义词。

[来源：GB/T 35248-2017，定义2.26]

##

安全 safety

免除了不可接受的风险的状态。

[来源：GB/T 35248-2017，定义2.24]

##

预期使用 intended use

按产品和/或系统提供的信息使用，无此类信息时，按通常理解的模式使用。

[来源：GB/T 29289-2012，定义3.2]

##

最低合理可行 as low as reasonably practicable

在可容许风险范围内，继续实施降低风险的措施，直到进一步降低风险的成本与预期实现的降低风险效果不相称为止。

# 缩略语

下列缩略语适用于本文件。

AI：人工智能（Artificial Intelligence）

# 智能功能及安全分类

## 智能功能

智能功能包括：

1. 感知，接收并转换信息的能力和过程；
2. 决策，对输入的信息进行处理并作出判断与决定的能力和过程；
3. 执行，将决策结果付诸实施的能力和过程；
4. 学习，吸收知识、经验、教训，以实现自适应、自调节的能力和过程。

## 安全分类

在合理可预见使用情况下，智能功能应满足以下要求：

1. 电气安全，在正常使用情况下，电气系统和机械结构不存在导致智能功能异常的不可接受的风险；
2. 功能安全，不存在由电气/电子系统的软件错误、硬件失效等功能异常影响智能功能而导致不合理的风险；
3. 信息安全，智能功能实现的过程中，不存在因敏感数据泄露、指令与数据篡改、服务被恶意中断等导致的不可接受的风险；
4. AI安全，AI系统在训练、部署和运行过程中，符合预期目标与伦理准则，具有较好的可解释性，不存在因功能不足、算法缺陷、数据偏差、人员误用等引发的不可接受的风险或违规行为。

# 安全设计总体原则

对智能产品智能功能的安全设计，宜在对各安全和危害类型深入分析的基础上，采用逐步迭代的思路进行，从而使智能产品智能功能的风险处于最低合理可行水平，并形成智能产品的智能功能设计方案。

在进行智能功能的安全设计过程中，除一般的安全设计要点外，还需注意以下内容：

1. 智能功能的安全设计与通常工业设计并行；
2. 考虑由智能功能带来的、一般产品不具备的特殊风险；
3. 考虑非智能产品为智能化而改变结构、材料等带来的风险。

# 安全设计流程

图1描述了智能功能的安全设计流程，包括“设计前准备”、“危害识别”、“安全设计”和“测试、评估与确认”，如经测试发现智能产品智能功能的风险未降低至最低合理可行水平，则需重新进行上述安全设计流程，直至智能产品智能功能的风险降至最低合理可行水平。智能产品智能功能的风险达到最低合理可行水平后，还需形成最终的智能产品智能功能设计方案。



图1 安全设计流程示意图

# 设计前准备

## 信息收集

对于智能产品，信息采集途径主要包括：

1. 相关法律法规，包括与智能产品相关的法律、行政法规、部门规章、地方性法规、地方规章等；
2. 相关产品标准，包括与智能产品相关的国家标准、行业标准、地方标准、国际标准和国外标准等；
3. 同类智能产品的事故报告；
4. 同类智能产品的召回信息；
5. 同类智能产品的检验、检测数据；
6. 同类智能产品的消费者投诉；
7. 专家意见；
8. 媒体报道；
9. 其他。

## 预期使用分析

### 使用人群分析

应确定智能产品的预期使用人群，宜重点考虑以下群体的智能功能预期使用情况：

1. 对智能技术具有较低熟练度的用户，如技术新手、儿童和老年用户等；
2. 对数据隐私风险具有较高敏感度的用户，如对机密性要求较高的职业人群等；
3. 有特殊需求的用户，如视觉或听力障碍者、肢体残疾者等；
4. 对兼容性要求较高的用户，如多平台、多设备用户等；
5. 其他用户。

应对用户的以下使用行为进行分析：

1. 正常操作行为，如遵循使用说明、适当配置和维护软硬件等；
2. 误操作行为，如错误配置、忽视警告、错误输入和不当环境使用等；
3. 故障下行为，如尝试故障排除、忽视故障问题和联系技术支持等。

### 产品特性分析

1. 应对智能产品的智能功能在特定环境下能达到的作用或效果进行分析；
2. 应对智能产品具有的智能功能安全问题，如电气安全问题、功能安全问题、信息安全问题、AI安全问题以及其他在实现智能功能时存在的风险进行分析；
3. 应对智能产品的智能功能承受黑客攻击、硬件破坏等外部恶意攻击行为的能力进行分析。

### 使用环境分析

宜对智能功能可能的使用环境，如通信网络、自然环境、交互控制、电磁环境等进行识别。除考虑正常的使用环境和可合理预见的非正常使用环境外，还宜考虑复杂、严酷的使用环境以及使用环境的突发情况。

# 危害识别

## 危害场景分析

基于智能功能的预期用途和使用场景，进行危害辨识：

1. 应通过实验模拟、行为观察等方式，模拟、发现用户在不同使用环境下使用智能产品的智能功能导致的不同危害发生场景，对上述场景中涉及的“用户-产品-环境”的关联关系进行致害机理分析，并分析在特定的危害场景下，由于使用智能功能可能导致的危害后果；
2. 宜重点考虑在各使用场景下应用智能功能可能使智能产品产生的新的风险和安全隐患。

## 危害识别方法

宜依据智能产品的安全特征、采集的风险信息、预期使用分析的结果，对风险源进行识别，危害识别方法包括但不限于：

1. 头脑风暴法及结构化访谈，一种收集各种观点及评价并将其在团队内进行评级的方法，头脑风暴法可以由提示、一对一以及一对多的访谈技术所激发；
2. 德尔菲法，一种综合各类专家观点并促其一致的方法，这些观点有利于支持风险源及影响的识别、可能性与后果分析以及风险评价，需要独立分析和专家投票；
3. 情景分析，在想象和推测的基础上，对可能发生的未来情景加以描述，可以通过正式或非正式的、定性或定量的手段进行情景分析；
4. 检查表，检查表提供了一系列典型的需要考虑的不确定性因素，使用者可参照以前的风险清单、规定或标准；
5. 预先危险分析（PHA），一种归纳分析方法，其目标是识别风险以及可能危害特定活动、设备或系统的危险性情况及事项；
6. 失效模式与影响分析（FMEA），一种识别失效模式、机制及其影响的技术，应在故障模式被纳入产品之前进行，进行FMEA分析需要确定关键件和重要件，实施FMEA需要掌握必要的信息和资料；
7. 危险与可操作性分析（HAZOP），一种综合性的风险识别过程，用于明确可能偏离预期绩效的偏差，并可评估偏离的危害度；
8. 危害分析与关键控制点（HACCP），一种系统的、前瞻性及预防性的技术，通过测量并监控那些应处于规定限值内的具体特征来确保产品质量、可靠性以及过程的安全性；
9. 系统理论过程分析（STPA），一种系统理论的过程分析方法，通过明确系统目的和事故，识别不安全控制行为，再基于致因场景分析制定控制措施。

## 典型危害类型

智能产品中与智能功能相关的危害类型包括：

1. 电气危害，由于智能产品的电气系统和机械结构影响智能功能实现而导致的危害；
2. 功能危害，因电气/电子系统的软件错误、硬件失效等功能异常引致智能功能不能正常实现导致的危害；
3. 信息安全危害，智能产品或功能因网络攻击或数据泄露而可能面临的暴露或损失；
4. AI危害，因智能产品的AI系统未达到预期目标或伦理准则，或可解释性较差导致的危害。

智能产品常见的危害见附录A。

# 安全设计

## 概述

针对已识别的危害，应通过智能产品安全设计，例如设计保护性措施、提供使用帮助和警示信息等减小其风险，使其达到最低合理可行水平。

## 电气安全设计

宜参照GB/T 25295规定的电气安全设计的原则和要求，对智能产品输、配、贮存、测量、监督、控制、调节、转换和消费电能的部件进行设计，并进行规定的检验或试验。

宜参照GB/T 15706规定的机械安全设计的原则和方法，对智能产品的机械零部件及其组合进行设计，并进行规定的检验或试验。

宜根据各类智能产品对应的电气安全标准进行电气安全设计，如：

1. 冰箱、空调、洗衣机、微波炉、烤箱、窗帘驱动器等功能型家电，参考GB/T 4706系列；
2. 电视机、音箱、计算机、显示器、打印机、数码相机、投影仪、手机等内容型家电，参考GB 4943.1；
3. 灯具，参考GB/T 7000系列等；
4. 开关、插座等电工设施，参考GB/T 16915系列、GB/T 2099系列和GB/T 1002等。

如果是多类设备交叉融合的智能产品，应考虑综合运用多个标准来评估电气安全；同时，还应注意标准之间存在的差异部分，应根据实际的工作环境选择最不利的标准要求进行考核。

## 功能安全设计

宜参照GB/T 41295.1规定的内容，基于智能功能的预期用途和工作环境对失效可能造成的危害情况进行充分的危害辨识和需求分析。

应参照GB/T 41295.2规定的内容制定智能功能的系统设计要求规范，通过系统的架构设计、系统的详细设计、软件设计、系统集成等完成功能安全设计，以确保智能产品的智能功能不存在因系统性失效和随机硬件失效而危及用户和环境的安全风险。

功能安全设计还宜遵循以下要求：

1. 对智能产品启动前系统运行状态进行判断，无失效/故障方可激活；
2. 采用机械结构部件作为最终的安全防护装置，确保智能产品的电子控制部件故障或失效时，能通过机械结构部件实现紧急处置和干预，避免引发安全事故；
3. 防护装置因检测到危险发出停止指令后，智能产品需维持停止状态，直至有安全重启条件为止；
4. 安全重启时，需先复位防护装置、解除停止指令，再恢复安全功能；
5. 对智能产品启动前和运行过程中可预见的安全风险进行预判和保护；
6. 正确应对无效命令（包括错误顺序的命令、未知命令、错误模式下的命令、错误的命令参数等），智能产品不受其影响正常运行；
7. 判断指令导致安全问题的可能性，对于可能导致安全问题的指令给出警告提示；
8. 具备保存日志记录的能力，以供智能产品问题的排查与修复；
9. 对于可远程操作的智能产品，在开始运行前，先在智能产品上取得操作授权，并获取智能产品当前的状态，在取得操作授权且智能产品处于安全状态时，才能进行远程操作。

## 信息安全设计

依据智能产品的信息产生、传递、接收、使用各阶段，对智能产品终端、控制端、平台等进行系统性设计，以确保智能产品不会在运行或受到攻击时产生数据泄露等信息危害，具体包括：

1. 应参照GB/T 35273和GB/T 32921的相关规定对用户的个人信息进行收集、存储和使用；
2. 应从硬件、固件、操作系统、应用、接口等多个方面确保信息的完整性、保密性、可用性；
3. 宜参照相关标准对智能产品标识、地理信息、通信安全、数据安全、密码功能和审计日志进行管理；
4. 对于可运行应用软件的智能产品，具备应用软件安装与更新的校验、对比与提示功能，且应用软件的安装与更新应经过授权用户的同意和确认，软件更新应确保更新时智能产品处于安全状态，且在发生预期外故障时能够回滚至更新前状态；
5. 对于带操作系统的智能产品，具备登录验证与身份鉴别功能。

## AI安全设计

基于智能功能的预期用途和使用场景，制定智能功能的系统设计要求规范，通过系统的规范定义和设计、系统的详细设计以确保智能产品智能功能不存在因预期功能不足引起的危害而导致的不合理风险，具体包括：

1. 宜采用多传感器冗余融合​，确保极端、边缘场景下的感知可靠性，避免因感知系统性能不足所导致不合理风险；
2. 宜采用对抗训练与模型冗余设计，通过对抗样本检测、鲁棒性测试等技术提升模型抗干扰能力，通过迭代优化算法，直到智能功能满足相应的风险可接受准则以最大限度地减少安全风险，确保输入扰动下的输出稳定性与可靠性，避免因算法模型设计不足导致不合理风险；
3. 宜考虑智能功能的可解释性设计，宜对高风险决策类AI功能提供决策依据的可解释性说明，确保关键逻辑可追溯；
4. 应考虑智能功能的伦理安全及合规性设计，避免算法设计中的偏见、歧视或违反社会公序良俗；
5. 宜采用基于角色的访问控制（RBAC）的权限分层与误用场景模拟训练，建立合理可靠的用户告知机制，确保用户操作合规性，避免合理可预见用户误操作。
6. 宜建立数据质量评估机制，避免因数据偏差或污染导致模型输出错误或歧视性结果。

# 测试、评估与确认

## 测试验证

宜邀请非研发团队的、有资质的第三方对安全设计流程和结果进行评审。宜在设计的适当阶段开展测试，测试可以在研发团队内部进行，也可以考虑通过外部第三方实验室开展；宜采用附录B推荐的方法导出测试策略，测试内容不宜局限于危害识别结果和安全设计效果，还应考虑因安全设计产生的新风险；宜针对测试工作制定计划，并将测试过程和结论文档化。测试过程的文档宜包括：

1. 测试前编制的测试规范，宜详细规定测试内容、方法、所用设备、步骤、预期结果等；
2. 测试过程中的测试记录；
3. 测试完成后编制的测试报告。

测试所用设备应具有完备的设备管理制度，设备执行测试前，应进行校准和功能完备性检查。宜采用自动化测试工具，减少手动操作的测试步骤。

## 风险评估

**11.2.1** 宜参照GB/T 27921规定的风险评估技术，结合信息收集、预期使用分析和测试验证结果，对每种危害的后果和可能性进行分析评估，度量每种危害的等级，以判断其是否在可容许范围内。

**11.2.2** 对于不可容许风险，应重新进行安全设计，确保其降至可容许风险水平；对于可容许风险，宜根据最低合理可行原则，进行进一步安全设计，直至继续降低风险的成本与预期降低风险的效果不相称（即达到最低合理可行水平）为止。

**11.2.3** 如有因安全设计产生的新危害，需对新危害一并进行评估。

**11.2.4** 对智能功能的风险进行评估，宜注意以下内容：

1. 与非智能功能危害之间存在因果、关联关系的智能功能危害，需考虑作为诱因的非智能功能危害发生的概率，或衍生危害的严重程度；
2. 结合危害场景的分析，需同时考虑初始场景发生的概率和危害场景中后果事件发生的概率；

## 设计方案确定

通过风险评估，确认产品智能功能的风险处于最低合理可行水平后，应根据相关设计内容和过程文档，形成最终的智能功能安全设计方案，并由相关方对安全设计方案进行确认。

附录A

（资料性）

智能产品危害示例

智能产品典型危害示例见表A.1。

表A.1 智能产品典型危害示例

|  |  |  |  |
| --- | --- | --- | --- |
| **危害类型** | **致害部件** | **危害** | **伤害或后果** |
| 电气危害 | 导线、电路板 | 短路 | 烧伤（火灾） |
| 电气危害 | 电池、电机 | 接触不良、散热不良 | 烧伤（火灾） |
| 电气危害 | 电机等驱动组件 | 旋转部件堵转 | 挤压伤、挫伤、绞伤 |
| 电气危害 | 导线 | 漏电 | 电击伤害 |
| 电气危害 | 电池、充电座 | 充放电异常 | 烧伤（火灾） |
| 功能危害 | 温控传感器 | 信号采集失效 | 烧伤（火灾）、冻伤 |
| 功能危害 | 激光雷达 | 点云采集失效 | 感知错误导致功能失控 |
| 功能危害 | 摄像头 | 图像采集失效 | 感知错误导致功能失控 |
| 功能危害 | 毫米波雷达 | 点云采集失效 | 感知错误导致功能失控 |
| 功能危害 | 控制逻辑单元 | 逻辑错误 | 产品失控 |
| 功能危害 | 人机交互界面 | 显示错误、显示异常 | 诱使用户误操作 |
| 功能危害 | 运动执行器 | 速度、方向等运动部件失效 | 运动功能失控 |
| 功能危害 | 智能系统内部接口 | 通信失效 | 产品失控 |
| 信息安全危害 | 无线通信模块 | 数据泄露 | 账户盗用、隐私泄露 |
| 信息安全危害 | 身份认证系统 | 伪造授权 | 非法入侵 |
| 信息安全危害 | 远程控制接口 | 指令劫持 | 设备恶意运行 |
| 信息安全危害 | 固件 | 固件更新文件遭篡改 | 信息泄漏、通信失效 |
| 信息安全危害 | 操作系统 | 系统权限设置错误 | 信息泄漏、通信失效 |
| AI危害 | 训练数据集 | 模型偏见 | 模型决策偏误 |
| AI危害 | 生成式模型 | 模型幻觉 | 产品失控 |
| AI危害 | 决策算法 | 逻辑失控 | 产品失控 |
| AI危害 | 摄像头 | 眩光抑制不足 | 模型误判导致产品失控 |
| AI危害 | 语音识别模块 | 噪音环境下的误触发 | 产品意外启动 |
| AI危害 | 多模态融合算法 | 跨模态信息同步偏差导致指令矛盾 | 产品失控 |
| AI危害 | 通信模块 | 恶意节点数据投毒 | 模型决策偏误 |
| AI危害 | 训练数据集 | 特殊环境下生物特征识别能力不足 | 产品不能正常启动 |

附录B

（资料性）

验证策略导出方法

考虑到智能功能的功能特性，验证策略（例如，集成测试用例、分析）的规范可使用适当的方法组合导出，见表B.1。

表B.1 验证策略的导出方法

| **编号** | **方法** |
| --- | --- |
| A | 边界分析 |
| B | 高加速寿命试验 |
| C | 性能测试 |
| D | 可靠性测试 |
| E | 故障插入测试 |
| F | 电磁兼容测试 |
| G | 气候试验 |
| H | 机械试验 |
| I | 等价类的生成与分析 |
| I | 边界值分析 |
| K | 错误推测法 |
| L | 判定表驱动分析法 |
| M | 功能图分析法 |
| N | 场景化端到端测试 |
| O | 偏见检测 |
| P | 可解释性分析 |
| Q | 数据驱动测试 |
| R | 对抗测试 |
| S | 渗透测试 |
| T | 模糊测试 |
| U | 安全功能验证测试 |
| V | 通信协议安全测试 |
| X | 数据匿名化测试 |
| Y | 操作系统安全测试 |

附录C

（资料性）

智能功能安全设计案例

## C.1 信息收集

A产品是市场上流程的一款智能门锁产品，具备密码、蓝牙、人脸、指纹开锁和认证信息修改等功能。通过文献检索、舆情搜索、实验检测、专家访谈等方法，采集与智能门锁相关的法律法规、标准、事故报告、召回信息、检验检测数据、消费者投诉、专家意见、媒体报道等渠道的信息。

## C.2 预期使用分析

A产品预期使用分析示例见表C.1。

表C.1 A产品预期使用分析

|  |  |
| --- | --- |
| 分析角度 | 具体描述 |
| 使用人群分析 | 重点考虑以下人群的使用：（1）年龄在8岁以下的未成年人，对智能技术具有较低的认知和熟练度；（2）在密码、指纹开锁功能使用中存在障碍的老年人或肢体残疾者。 |
| 产品特性分析 | 产品作用和效果：（1）密码开锁，通过一次性、临时或永久密码进行开锁；（2）蓝牙开锁，手机控制端通过蓝牙功能与智能门锁产品终端进行连接，向产品传输密码实现开锁；（3）生物特征开锁，通过指纹或人脸比对进行开锁；（4）认证信息修改，在产品终端或控制端对产品的认证信息进行修改。产品智能功能安全问题：（1）电气安全问题，如锁舌卡死时持续向电机供电引致电机烧毁、引燃周边可燃物，诱发火灾；（2）功能安全问题，如断电引致的门锁无法正常开启；（3）信息安全问题，如数据存储不当导致的用户生物特征泄露；（4）AI安全问题，如活体检测能力不足引致的门锁被异常开启。 |
| 使用环境分析 | 可适应短时间的高温或低温且不影响使用，可适应一定强度的正弦振动、冲击和自由跌落且不影响使用，具备一定的电磁兼容性。 |

## C.3 危害识别

### C.3.1 危害场景构建与分析

场景1：因长期使用，锁舌部分机械结构异常导致卡死，同时电源持续向电机进行供电，引致电机过热、烧毁，甚至引燃周边可燃物诱发火灾。

场景2：因断电或突发事件，产品无法正常从电网获得供电，无法实现密码或生物特征的认证，导致门锁无法正常开启。

场景3：因权限设置错误，外部用户可以使用一次性或临时密码获取、修改产品的认证信息，导致门锁防盗功能失效、用户生物特征泄露。

场景4：因活体检测能力不足，外部用户可以通过复制指纹伪造、人脸照片伪造等实现门锁的异常开启。

### C.3.2 识别方法选择

采用德尔菲法和情景分析法开展危害识别。

### C.3.3 危害类型确定

经来自企业、科研机构、监管部门的7名专家初判，并通过对危害场景的情景分析，上述危害场景的识别结果见表C.2。

表C.2 危害类型识别结果示例

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **场景** | **危害类型** | **致害部件** | **危害** | **伤害或后果** |
| **场景1** | 电气危害 | 电机等驱动组件 | 旋转部件堵转 | 失火 |
| **场景2** | 功能危害 | 智能系统 | 功能失效 | 产品功能失效 |
| **场景3** | 信息安全危害 | 操作系统 | 权限设置错误 | 信息泄漏、产品功能失效 |
| **场景4** | AI危害 | 训练数据集 | 模型决策偏误 | 产品功能失效 |

## C.4 安全设计

针对场景1，拟采用如下安全设计：使用符合要求的材料和工艺设计电气部件和机械部件，使之具有较好的耐久性；部署监测系统监测锁具启、闭操作执行情况，操作执行异常时自动终止操作。

针对场景2，拟采用如下安全设计：装备应急机械防盗锁头，使产品在断电情况下能正常启、闭；使用可充电电池作为供电电源。

针对场景3，拟采用如下安全设计：进行合理的权限控制，遵循最小必要原则，以防止一次性或临时密码能够获取、修改产品的认证信息；在对认证数据进行获取、修改时，采用双重验证机制以保护数据的机密性。

针对场景4，拟采用如下安全设计：合理设置训练模型所用数据集，使模型具备活体检测能力。

## C.5 评估与确认

### C.5.1 测试验证

针对场景1，拟通过如下方法进行测试验证：在额定电压和额定负载电流情况下，进行10000次锁具启、闭操作，试验后不应有电气部件或机械部件的损坏或失效，且应能正常工作；限制锁具启、闭操作部件，测试锁具自动终止功能能否正常运行且达到预期效果。

针对场景2，拟通过如下方法进行测试验证：在不放入电池、不接通电源的情况下，使用机械钥匙对受试产品进行操作，确保其能够正常启、闭；在电池满电、不接通电源的情况下，对受试产品连续进行启、闭操作，确保操作能连续进行3000次以上。

针对场景3，拟通过如下方法进行测试验证：设计测试类测试不同类型用户对认证信息的获取和修改权限，确保认证信息的使用遵循最小必要原则，且在获取和修改时双重验证机制有效。

针对场景4，拟通过如下方法进行测试验证：使用生物特征的复制、照片、假体、AI深度合成、视频等伪造对受试产品进行测试，确保受试产品能检测或防止伪造数据的使用。

### C.5.2 风险评估

根据测试结果对产品安全性进行评价，若产品安全性未达到最低合理可行要求，则重新进行安全设计，直至其安全性达到最低合理可行要求。

### C.5.3 设计方案确定

重复上述过程，直至产品安全性达到最低合理可行要求，之后形成A产品最终的智能功能安全设计方案，并指导A产品的工业生产。

参考文献

1. GB/T 20002.4-2015 标准中特定内容的起草 第4部分：标准中涉及安全的内容
2. GB/T 20438.1-2017 电气/电子/可编程电子安全相关系统的功能安全 第1部分：一般要求
3. GB/T 20438.5-2017 电气/电子/可编程电子安全相关系统的功能安全 第5部分：确定安全完整性等级的方法示例
4. GB/T 21109.3-2007 过程工业领域安全仪表系统的功能安全 第3部分：确定要求的安全完整性等级的指南
5. GB/T 28219-2018 智能家用电器通用技术要求
6. GB/T 29289-2012 消费品安全设计通则
7. GB/T 34400-2017 消费品召回 生产者指南
8. GB/T 35248-2017 消费品安全 供应商指南
9. GB/T 40979-2021 智能家用电器个人信息保护要求和测评方法
10. GB/T 41295.3-2022 功能安全应用指南 第3部分：测试验证
11. GB/T 41302-2022 工业产品数据字典通用要求
12. GB/T 41387-2022 信息安全技术 智能家居通用安全规范
13. GB/T 41789-2022 智能家用电器的通用安全技术要求
14. GB/T 41790-2022 智能消费品质量安全改进指南
15. GB/T 43942-2024 智能船舶风险评估方法
16. GB/T 44602-2024 网络安全技术 智能门锁网络安全技术规范
17. GB/T 45097.1-2024 智能消费品安全 第1部分：危害（源）识别
18. GB/T 45097.2-2024 智能消费品安全 第2部分：风险评估
19. GA 374-2019 电子防盗锁