

# 团 体 标 准

T/ISC 0004—2020

---

## 网络运营者针对未成年人的有害信息 防治体系建设框架

Framework for network operators of protecting minors from harmful content

2020-09-24 发布

2020-12-01 实施

---

中 国 互 联 网 协 会 发 布



## 目 次

前言 .....	III
引言 .....	IV
1 范围 .....	1
2 规范性引用文件 .....	1
3 术语和定义 .....	1
4 缩略语 .....	1
5 参考框架 .....	1
5.1 框架概述 .....	1
5.2 防治技术 .....	2
5.2.1 概述 .....	2
5.2.2 安全技术 .....	2
5.2.3 内容分级 .....	3
5.2.4 协同治理 .....	3
5.3 防治手段 .....	3
5.3.1 预防 .....	3
5.3.2 防护 .....	4
5.3.3 治理 .....	5
5.4 防治保障 .....	5
5.4.1 资源保障 .....	5
5.4.2 人才培养 .....	5
5.4.3 标准规范 .....	5
5.4.4 合规运营 .....	6
5.4.5 安全检测 .....	6
附录 A（资料性） 针对未成年人的有害信息防治相关方及工作维度 .....	7
附录 B（资料性） 网络有害信息对未成年人的影响分析维度 .....	8
附录 C（资料性） 有害信息的防治实践案例 .....	9
附录 D（资料性） 有害信息举报流程示例 .....	10
附录 E（资料性） 内容涉及未成年人的网络色情信息分类示例 .....	11



## 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。

本文件由中国互联网协会提出并归口。

本文件起草单位：深圳市腾讯计算机系统有限公司、国家计算机网络应急技术处理协调中心、中国标准化研究院、作业帮教育科技(北京)有限公司、百度在线网络技术（北京）有限公司、深圳市网安计算机安全检测技术有限公司、合肥天帷信息安全技术有限公司、北京网络行业协会、恺英网络股份有限公司、今日头条有限公司。

本文件主要起草人：黄超、赵芸伟、章建方、张文彬、刘广、洪跃腾、武建双、王琮玮、苏文颖、杨晓栋、王笛、施秀云、罗雯、李笑如、欧瑜、肖薇、何欣、赵玉现、盛安宇、张娜、刘洋、袁若熙、舒梦莹、庞永杰、陈晓冉、钟振坤、韩晗、代威。

## 引 言

在网络和信息技术高速发展的浪潮下，我国未成年人触网率及未成年人网民数量迅速上升，网络有害信息对未成年人健康成长造成负面影响的事件频有发生。加强未成年人网络保护尤其是对有害信息的防护和治理成为社会共识。

我国已有相关法律法规明确提出保护未成年人不受有害信息侵害的要求。《中华人民共和国网络安全法》提出国家支持研究开发有利于未成年人健康成长的网络产品和服务，依法惩治利用网络从事危害未成年人身心健康的活动，为未成年人提供安全、健康的网络环境。《中华人民共和国未成年人保护法》（修订草案）提出任何组织和个人不得利用网络制作、复制、发布、传播含有淫秽、色情、暴力、邪教、迷信、凶杀、恐怖、赌博、涉毒等法律、行政法规禁止内容的信息。因此有必要进一步加强引导和规范网络运营者针对未成年人的有害信息防治工作。

# 网络运营者针对未成年人的有害信息防治体系建设框架

## 1 范围

本文件确立了网络运营者针对未成年人的有害信息防治体系建设框架，给出了有害信息防治手段、防治保障、防治技术等内容。

本文件适用于网络运营者，为其开展未成年人网络有害信息防治工作提供参考和指引。

## 2 规范性引用文件

本文件没有规范性引用文件。

## 3 术语和定义

下列术语和定义适用于本文件。

### 3.1

**未成年人** minors

未满十八周岁的公民。

### 3.2

**有害信息** harmful content

网络中涉及到未成年人以及危害未成年人身心健康的违反我国法律法规的违法信息和不良信息。

### 3.3

**网络运营者** network operators

网络的所有者、管理者和网络服务提供者。

注：参照《中华人民共和国网络安全法》第七十六条中网络运营者的定义。

## 4 缩略语

下列缩略语适用于本文件：

AI (Artificial Intelligence) 人工智能

OCR (Optical Character Recognition) 光学字符识别

## 5 参考框架

### 5.1 框架概述

网络运营者针对未成年人有害信息的防治体系建设框架应分为防治技术、防治手段、防治保障三个方面。参考框架见图1，其中，防治技术是为实现防治手段所需要建设和加强的共性技术和基础能力；防治手段为具体防治措施，提供实现未成年人网络有害信息防治工作的直接方法、工具和能力，包括预

防、防护、治理三个层面，涵盖有害信息防治事前、事中、事后全过程；防治保障是开展未成年人网络有害信息防治工作的保障措施包括资源保障、人才培养、标准规范、合规运营和安全检测。

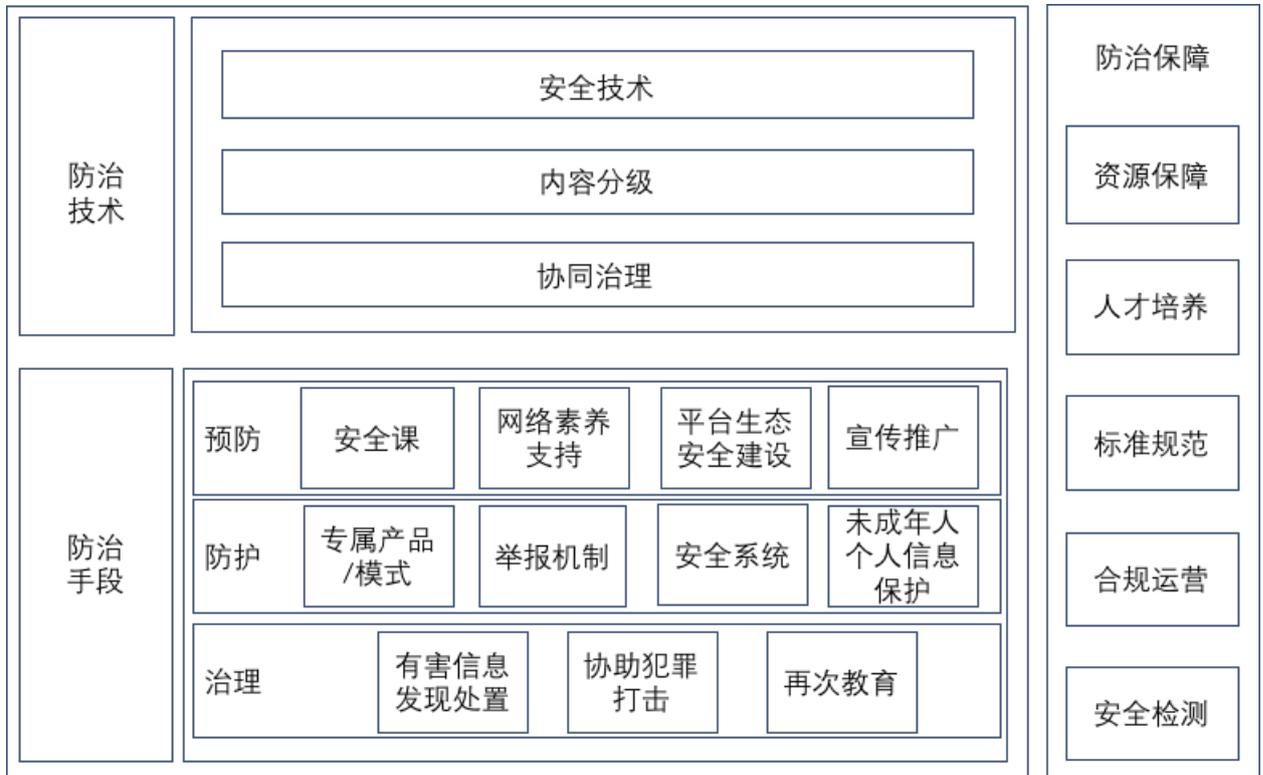


图1 有害信息防治体系建设框架

## 5.2 防治技术

### 5.2.1 概述

网络运营者应认识到针对未成年人的网络有害信息防治工作是一项综合、复杂的工作，面临对抗技术不断升级、防护技术需不断演进等技术挑战，需要联合相关方进一步加强研究，综合运用多种技术手段实现保护未成年人免受有害信息侵害的安全目标。

有害信息防治工作相关方可参见附录 A。

### 5.2.2 安全技术

本文件所指安全技术是与有害信息的安全防护和治理相关的防治技术，包括有害信息发现技术、网络行为识别技术、网络信息处理基础技术、基础库建设。网络运营者宜加强以下技术研发和实践工作：

- a) 有害信息发现技术：
  - 1) 多媒体（视频、音频、图片、文字等）内容安全识别技术；
  - 2) 基于 AI 的内容分类、有害信息识别等算法的设计和模型训练，如深度学习算法等；
  - 3) 基于 OCR 的图片内容识别技术；
  - 4) 基于相似度的图片识别技术等。
- b) 网络行为识别技术：用于识别违法用户行为，如恶意注册、恶意发布、分享有害信息等恶意行为。

- c) 网络信息处理基础技术：
  - 1) 数据采集和预处理技术；
  - 2) 高速网络流量处理技术，如网络协议解析；
  - 3) 大数据存储、查询技术等。
- d) 基础库建设：
  - 1) 样本库：有害词库、语料库、视频库、图片库、有害外部链接库、有害 IP 地址库等，支持样本的新增、修改、删除等样本库管理能力；
  - 2) 特征库：对提取的有害样本、有害行为等特征的存储和管理。

### 5.2.3 内容分级

网络运营者应积极研究针对未成年人的内容分级管理机制，研究适宜未成年人接触，与其年龄、身份相符的网络信息筛选标准，例如把网络信息分为适宜未成年人观看和不宜未成年人观看两种级别。

有害信息对未成年人的侵害判定维度可参见附录 B。

### 5.2.4 协同治理

网络运营者之间应加强协同和配合，从技术协同、数据协同出发，逐步实现更大范围的组织协同、机制协同的协同治理工作体系。

## 5.3 防治手段

### 5.3.1 预防

#### 5.3.1.1 安全课

- a) 网络运营者宜开设针对未成年人的网络安全课程，引导和教育未成年人正确认识和使用网络，提高未成年人识别网络有害信息的能力和使用未成年人保护相关工具/技术的水平，促进未成年人树立正确的网络观，自觉抵制有害信息；
- b) 网络运营者宜开设针对监护人、教师等与未成年人有害信息防治工作重要相关方的网络安全课程，供其了解和掌握如何关注未成年人网络行为，并从自身做起，以身作则，引导未成年人规范上网行为；
- c) 网络运营者可联合政府、互联网企业或行业组织、科研机构等加强对网络安全课程的开发，以及增加对未成年人网络心理研究等方面的投入，丰富和提升安全课的理论基础、教学形式和宣传推广模式等。

未成年人网络安全课实践可参见附录 C。

#### 5.3.1.2 网络素养支持

- a) 网络运营者宜联合政府或其他相关组织，加强未成年人、监护人、网民等应具备的网络素养（例如未成年人安全上网、女童网络素养等）研究，提升未成年人对网络信息健康程度辨别能力，促进未成年人增强自我保护意识和能力；
- b) 网络运营者应结合自身业务和技术特点，开发未成年人网络素养教育工具，例如音视频教育产品、辅导手册、教育功能游戏等。

网络素养教育支持实践可参见附录 C。

#### 5.3.1.3 平台生态安全建设

- a) 网络运营者提供网络平台服务的，应加强本平台网络信息内容生态安全治理工作，制定本平台网络信息内容发布规则，为未成年人营造清朗的网络空间；
- b) 通过向平台上的信息内容生产者进行平台网络安全培训，为平台上的信息内容生产者提供内容安全服务（例如内容安全 API 服务等）等方式，提升平台上信息内容生产者的安全意识和技术能力，维护平台的信息内容生态安全。

平台生态安全建设实践可参见附录 C。

#### 5.3.1.4 宣传推广

网络运营者应利用其自身宣传渠道或联合传统媒体、自媒体等，加强网络有害信息防治有关法律、政策、方法、工具、产品规则等网络安全宣传和推广力度，提升未成年人网络保护各相关方的安全守法意识和能力，增强未成年人网络行为的自觉性和自我保护能力。

宣传推广实践可参见附录 C。

#### 5.3.2 防护

##### 5.3.2.1 未成年人专属产品/模式

###### a) 专属产品

针对未成年人的安全需求，网络运营者应针对性地研发专门的未成年人专属网站、专栏、专区、产品或服务，提供教育类、知识类等适宜未成年人接触的网络内容。

###### b) 专属模式

在长视频、短视频、直播等视频和直播类产品和服务中，网络运营者应加大对未成年专属模式的研制，提供如“青少年模式”等未成年人专属使用模式，且该模式支持：

- 1) 专门为未成年人精选推荐适合未成年人的网络内容；
- 2) 支持限制/取消部分功能（如充值、打赏等）的使用；
- 3) 支持使用时长、使用时间段等限制。

##### 5.3.2.2 举报机制

网络运营者应建立举报机制，包括：

- a) 设立相应的有害信息投诉/举报入口和举报流程（有条件的网络运营者还可在提供的产品/服务举报入口中设置侵害未成年人权益的专属举报标签）；
- b) 配备专门的举报处置人员；
- c) 及时受理和处置相关举报。

举报流程的设立可参见附录 D。

##### 5.3.2.3 安全系统

网络中每天都产生大量数据，从海量公开场景的数据中发现和处理有害信息依赖安全系统和工具的支持。网络运营者宜加强内容安全相关安全系统的建设，包括但不限于以下系统：

- a) 内容安全机器检测系统，提供通过基于规则检测或基于机器学习等方式，对信息内容进行自动化安全检测，用于提高检测效率，降低人工成本；
- b) 人工审核系统，提供审核操作平台，用于相关审核人员在平台上高效的审核各类视频、图像、文本等内容，提升审核效率；
- c) 样本数据标注系统，提供对获取的样本数据的标注能力，用于后续的模式训练等工作；
- d) 内容安全验证工具，提供对安全策略实施准确率、相关安全系统运行状态等有效性验证；

e) 内容安全统计分析系统，提供对网络运营者整体内容安全状态的分析、展示等能力。安全系统实践可参见附录 C。

#### 5.3.2.4 未成年人个人信息保护

网络运营者应加强对未成年人个人信息的保护：

- a) 收集和使用未成年人个人信息时，应经过监护人同意；
- b) 在处置相关安全事件时，可通过对需展示的个人信息采取去标识化处理、匿名化等方式，加强对未成年人个人信息的保护；
- c) 遵照法律法规的规定，严格限制共享、转移、披露未成年人个人信息的情形。

### 5.3.3 治理

#### 5.3.3.1 有害信息治理

网络运营者应通过以下方式加强有害信息治理工作：

- a) 建立有害信息治理相关的管理制度；
- b) 设立有害信息治理相关的管理机构和人员；
- c) 建设有害信息治理相关的安全系统，通过机器检查和人工审核相结合的方式及时发现和处置有害信息。具体处置方式包括信息清除、功能干预等。信息清除指通过删除、展示屏蔽等方式使有害信息无法被查阅；功能干预指对违规账户进行限制发布、限制提现等功能限制。具体处置方式的选择应根据网络运营者的业务特点确定。

#### 5.3.3.2 协助犯罪打击

对于发现制作、发布、传播网络有害信息等违法行为，网络运营者应当为公安机关、国家安全机关依法维护国家安全和侦查犯罪的活动提供技术支持和协助，提高打击犯罪的有效性，保护未成年人权益。

#### 5.3.3.3 再次教育

网络运营者应通过视频教学、案例分析、警示参观等方式，引导未成年人进行相关法律法规、知识的学习，达到再次教育的效果。

### 5.4 防治保障

#### 5.4.1 资源保障

网络有害信息的产生、传播、变种速度快，防治工作难度大。网络运营者应加强投入相关的资源，包括经费资源、人力资源、技术资源等，保障本组织有害信息防治工作可以长期有效开展。

#### 5.4.2 人才培养

实现有害信息的防治工作离不开专业人才的技能和经验。网络运营者应加强培养在法律、技术、教育、网络安全等专业上的复合型人才，提升本组织有害信息治理人员的专业水平。

#### 5.4.3 标准规范

网络运营者应加强参与针对未成年人的有害信息防治相关标准（如人才评定、安全技术和系统、未成年人专属产品/模式、安全测评验证等方向）的研制和应用推广工作，满足对人才培养、能力建设、有害信息分类分级、效果评价等有害信息防治工作的实际需求。

内容涉及未成年人的网络色情信息分类示例可参见附录 E。

#### 5.4.4 合规运营

网络运营者应加强对未成年人网络有害信息防治相关法律法规工作的支持，并在业务运营过程中做到合法合规。

#### 5.4.5 安全检测

网络运营者提供的面向未成年人的产品/服务上线发布前，应开展相关安全检测验证工作，保障在内容上不含有害信息、在功能上提供了相应安全保护能力后方可上线。

## 附录 A

(资料性)

## 针对未成年人的有害信息防治相关方及工作维度

针对未成年人的有害信息防治相关方及工作维度如下表A.1所示。

表A.1 相关方及工作维度

序号	相关方	工作维度
1	网络运营者	加强针对未成年人的有害信息防治工作，履行企业社会责任。
2	未成年人	坚决抵制网络有害信息，保证自身网络行为的合法合规。
3	监护人	依法履行对未成年人的监护职责，保障未成年人网络行为安全，保护未成年人健康成长。
4	学校和教师	对未成年人的教育过程以及未成年人在校管理过程进行网络安全教育和监督等方面发挥积极作用。
5	网民	保证自身网络行为的合法合规，积极发挥网民监督作用。
6	政府	为未成年人有害信息防治工作提供政策基础和保障，履行相关的监督管理职能，引导全社会为未成年人营造清朗健康的网络空间。
7	社会组织	积极开展有利于未成年人网络健康的各项社会活动和服务。 例如 科研机构开展未成年人网络心理、网络安全技术等研究工作；服务机构开展安全培训、安全咨询、安全测评等服务。

## 附录 B

(资料性)

### 网络有害信息对未成年人的影响分析维度

分析网络信息是否可能危害未成年人身心健康，可以从以下维度开展：

- a) 引诱未成年人犯罪。通过传播暴力血腥、赌博、制毒贩毒吸毒、制作售卖违禁品、恐怖活动等方式，直接或间接的诱导未成年人犯罪或实施违反法规的行为。
- b) 引诱未成年人自杀。通过宣传、介绍、鼓励自残、自杀等形式，直接或间接的引诱未成年人实施自杀行为。
- c) 影响未成年人性成熟和性心理健康。通过传播淫秽色情、低俗挑逗等信息，可能造成未成年人对色情活动的模仿，影响未成年人性心理、性生理、性道德和性法制观念的扭曲，导致犯罪等问题出现。
- d) 对未成年人进行网络欺诈。利用未成年人的猎奇心理以及不成熟的认知辨别能力等，对未成年人实施诈骗，或诱导未成年人对他人进行诈骗。
- e) 侵害未成年人隐私。通过诱骗、隐瞒、黑客等手段，非法搜集、非法使用、售卖、曝光未成年人的个人信息。
- f) 对未成年人进行网络欺凌。通过辱骂、诽谤、人肉、曝光个人信息、威胁恐吓、制造散播谣言等手段，影响到未成年人精神状态或现实生活。

## 附录 C (资料性) 有害信息的防治实践案例

### C.1 安全课实践

案例一：腾讯联合全国“扫黄打非”工作小组办公室开展以“护苗 网络安全进课堂”为代表的未成年人网络安全教育系列活动。案例介绍：

<http://www.shdf.gov.cn/shdf/contents/767/409224.html> 《“护苗·网络安全进课堂”2019乡村行活动》。

### C.2 网络素养教育支持实践

案例一：腾讯举报中心刊载《繁星特刊》系列公众号网络素养教育文章，保护未成年人上网安全。案例介绍：<https://mp.weixin.qq.com/s/UZFKLyN3q70Ms8Ap6qHq0w>《繁星特刊 | 未成年人上网避坑指南》。

### C.3 平台安全生态建设实践

案例一：腾讯公司举办阳光媒体人大会，为微信公众平台和小程序账号运营者以及直播、短视频机构代表开展网络内容安全运营规则培训、交流内容安全生态建设经验。

案例介绍：<https://new.qq.com/rain/a/20191231a0700300>《2019年度阳光媒体人大会》。

案例二：腾讯微信推出“珊瑚内容安全助手”小程序，帮助微信公众号运营者和小程序开发者通过小程序一键鉴别内容和产品的安全风险，在内容安全方面为运营者和开发者提供更多帮助。

案例介绍：<https://mp.weixin.qq.com/s/j8ZrlikmjgDLDxzMeP2NA>

《腾讯“珊瑚安全”API正式上线，小程序开发者可接入内容安全服务》

### C.4 宣传推广实践

案例一：腾讯游戏官网解读、宣传腾讯公司发起的国内互联网游戏行业首个面向未成年人健康上网的系统解决方案。

案例介绍：<https://games.qq.com/a/20170217/023632.htm>《腾讯游戏成长守护平台上线》

### C.5 安全系统支持实践

案例一：作业帮发布国内首个青少年信息安全综合防护平台，用大数据、AI技术守护“数字一代”网络安全，旨在为未成年人营造清朗的网络空间。

案例介绍：<https://baijiahao.baidu.com/s?id=1612855191532377343&wfr=spider&for=pc>《“寸功”初心：用大数据、AI技术守护“数字一代”网络安全》

附录 D  
(资料性)  
有害信息举报流程示例

有害信息举报流程可以包括信息举报、信息审核、处置归档、配合执法四个流程。具体如下：

- a) 信息举报。信息举报是网络有害信息的来源之一，该阶段用户通过网络平台的指引寻找到举报入口，根据举报指引，提交举报内容。有条件的平台可以研究设立针对未成年人的专属举报标签，对未成年人提供特殊保护。
- b) 信息审核。审核确定对举报信息进行审理鉴定，网络平台员工依据审核专业度，处理举报信息，过滤无效信息，惩罚恶意举报。
- c) 处置归档。网络平台对举报信息进行处置响应，同时将结果反馈给举报用户，并对举报处置相关信息进行归档。
- d) 配合执法。依据法律法规要求，配合相关执法机关工作。

附录 E  
(资料性)

内容涉及未成年人的网络色情信息分类示例

未成年人网络色情信息可大致分为一般色情信息和内容涉及未成年人的色情信息。

一般色情信息指不是专门针对未成年人或内容包含未成年人的色情信息，具体可参考《全国人大常委会关于惩治走私、制作、贩卖、传播淫秽物品的犯罪分子的决定》、《关于认定淫秽及色情出版物的暂行规定》等有关法律法规。

内容涉及未成年人的色情信息特指内容包含未成年人的色情信息。本附录为内容涉及未成年人的色情信息判别提供参考示例。

表E.1 内容涉及未成年人的色情信息分类示例

序号	分类	示例
1	性侵害	展示强奸、偷拍、性骚扰、性虐待、性侮辱未成年人等。
2	性行为	直接展示未成年人的呻吟、叫床等涉及性行为的内容等。
3	性交易	表现涉及未成年人的卖淫、嫖娼、援交的过程等。
4	性怪癖	表现恋童癖、涉及未成年人的乱伦等。
5	以未成年人为宣传点的色情推广	以未成年人为宣传点，推介淫秽色情网站、APP等。