

ICS 33.050

CCS M 30

团 体 标 准

T/TAF 138—2022



App 推荐算法用户权益保护技术要求及测 评规范

App user rights protection requirements and evaluation specification in
recommendation algorithm

2022-11-25 发布

2022-11-25 实施

电信终端产业协会 发布

目 次

前言	II
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 基本原则	2
6 个人信息保护要求	2
7 推荐算法技术安全要求	2
7.1 安全目标概述	2
7.2 算法模型安全	2
7.3 算法安全可控	3
7.4 算法应用安全	4
8 用户主体权益保护要求	5
8.1 一般用户主体权利	5
8.2 特殊群体权益	6
9 推荐算法安全管理要求	6
9.1 安全运营	7
9.2 安全评估	7
9.3 应急处置	7
9.4 违规追责	7
10 测试方法	7
附录 A（资料性）常见推荐算法应用	9
参考文献	10

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。

本文件中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。

本文件由电信终端产业协会提出并归口。

本文件起草单位：中国信息通信研究院、北京三快在线科技有限公司、泰尔认证中心有限公司、北京抖音信息服务有限公司、北京快手科技有限公司、OPPO广东移动通信有限公司、郑州信大捷安信息技术股份有限公司、荣耀终端有限公司、蚂蚁科技集团股份有限公司、咪咕互动娱乐有限公司。

本文件主要起草人：祖岩岩、刘瑾、杜云、宁华、黄坤、张波、方强、周飞、李京典、李可心、王高林、冷杉、吴斌、王芳、常琳、赵昕、王宇晓、李昶婧、王昕、付艳艳、刘为华、刘献伦、赵晓娜、杨骁涵、吴少卿、林冠辰、王宝刚。

App 推荐算法用户权益保护技术要求及测评规范

1 范围

本文件规定了App推荐算法在用户权益保护方面应满足的基本原则、个人信息保护要求、推荐算法技术安全要求、用户主体权益保护要求、推荐算法安全管理要求及测评方法。

本文件适用于App推荐算法服务提供者规范其个人信息处理活动、算法建模及使用行为，也适用于监管部门、第三方评估机构等组织对App推荐算法服务进行监督、管理和评估。

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/T 35273—2020 信息安全技术 个人信息安全规范

3 术语和定义

GB/T 35273—2020界定的以及下列术语和定义适用于本文件。

3.1

推荐算法 recommendation algorithm

利用生成合成类、个性化推送类、排序精选类、检索过滤类、调度决策类等算法技术向用户提供信息。

注：常见的推荐算法应用见附录A。

3.2

个性化推送 personalized push

基于对特定用户或用户群体进行个性化用户画像实现的信息推送，信息推送的展现形式包括但不限于推荐、列表页面、弹窗等。

3.3

用户标签 user tag

通过高度概括、容易理解的特征来描述用户在进行信息服务交互过程中形成的标识。

4 缩略语

下列缩略语适用于本文件。

App: 移动应用软件 (Mobile Application)

5 基本原则

App推荐算法服务提供者应满足以下原则。

- a) 合法合规: 应符合相关法律法规的规定, 避免推荐算法服务引起社会公平、道德伦理、个人信息安全、网络安全等方面的风险;
- b) 公平公正: 利用个人信息进行推荐算法服务的结果公平、公正, 不得实行不合理的差别待遇。不得利用推荐算法服务开展不正当竞争;
- c) 公开透明: 以适当方式向用户公开推荐算法服务处理活动处理个人信息的范围、目的、规则;
- d) 科学合理: 根据行业类型、业务场景等实际情况, 科学、合理地设置相关参数条件和权重、决策规则。

6 个人信息保护要求

App推荐算法服务如涉及收集、存储、使用、共享、转让、公开披露、删除等个人信息处理活动的, 除满足GB/T 35273的要求, 还应满足以下要求:

- a) 应以显著方式告知用户提供推荐算法服务的主要场景和类别, 如个性化推送的目的、使用个人信息的种类、用户画像的目的, 以及开启或关闭个性化推送的自主管理控制方式等内容;
- b) 除为达到个人信息主体授权同意的使用目的所必需外, 使用个人信息时应消除明确身份指向性, 避免精确定位到特定个人。例如, 为准确评价个人信用状况, 可使用直接用户画像, 而用于推送商业广告目的时, 则宜使用间接用户画像;

注: 根据GB/T 35273, 直接使用特定自然人的个人信息, 形成该自然人的特征模型, 称为直接用户画像。使用来源于特定自然人以外的个人信息, 如其所在群体的数据, 形成该自然人的特征模型, 称为间接用户画像。

- c) 利用个人信息进行推荐算法服务, 应事前开展个人信息保护影响评估并留存相关记录 (不低于3年), 评估内容包括但不限于:
 - 1) 个人信息的处理目的、处理方式等是否合法、正当、必要;
 - 2) 对个人权益的影响及安全风险;
 - 3) 所采取的保护措施是否合法、有效并与风险程度相适应。

7 推荐算法技术安全要求

7.1 安全目标概述

推荐算法技术安全的目标为推荐算法结果公平公正、科学合理, 主要要求包括: 算法模型安全、算法安全可控、算法应用安全。

7.2 算法模型安全

7.2.1 数据处理

App推荐算法服务提供者应满足以下要求:

- a) 算法模型如使用用户数据, 应对违法和不良信息进行过滤 (针对训练时除外);
- b) 采用机器学习算法时训练数据的处理要求如下:

- 1) 宜对训练数据进行预处理，包括数据清洗、数据脱敏、数据标注等；
- 2) 训练数据如涉及个人信息，应加密存储并应严格控制访问权限，具备对数据访问和操作等行为审计能力；
- 3) 宜使用不同于训练的独立数据集进行验证，并通过采样或者扩充样本等方式提高训练数据的多样性和公平性。

7.2.2 用户标签

App推荐算法服务提供者应满足以下要求：

- a) 用户标签涉及指向特定个人的敏感个人信息的，应建立个人信息的脱敏处理机制；
- b) 用户标签不应包含淫秽、色情、赌博、迷信、恐怖、暴力的内容或表达对民族、种族、宗教、残疾、疾病歧视的内容；
- c) 应建立用户标签审核细则，不得将违法和不良信息关键词作为用户标签并据以推送信息。

7.2.3 模型应用

App推荐算法服务提供者应满足以下要求：

- a) 宜采取加密、混淆等技术手段，提供对模型参数的安全保护手段；
- b) 应完善记入用户模型的兴趣点规则，不得将违法和不良信息关键词记入用户兴趣点并据以推送信息；
- c) 应定期审核、评估、验证算法服务机制机理，不得设置诱导用户沉迷、过度消费等违反法律法规或者违背伦理道德的算法模型。

7.3 算法安全可控

7.3.1 安全审核

App推荐算法服务提供者提供网络信息内容服务的，应满足以下要求：

- a) 宜建立内容审核团队并完善相关管理制度，包括组织架构、职责分工、团队规模、培训、考核等；
- b) 宜建立支撑审核人员高效审核的技术体系，包括审核平台、判研知识库、审核标准等；
- c) 宜建立不良信息特征库管理规范，包括入库标准、出库规则等。

7.3.2 干预控制

App推荐算法服务提供者应满足以下要求：

- a) 应建立排名干预能力，如排序公式引入人工权重项，包括触发条件、干预目的、干预手段等；
- b) 宜建立打榜行为发现能力，如数据趋势异常检测，防范恶意刷榜等行为；
- c) 应建立不良内容发现能力，如通过关键字、图像识别模型和规则模板识别不良文字和图片信息。

7.3.3 风险监测

宜对算法应用过程中的风险点（不符合规范预期的结果，如不良类目、错误结果、异常结果等）进行监测，并对结果进行多维研判和精准处置。

7.3.4 追溯机制

App推荐算法服务提供者应满足以下要求：

- a) 数据处理、标签管理、模型训练和模型推理等算法服务过程涉及人员，应对相关设计、操作进行登记记录，以明确主体责任、严格责任制度；

- b) 训练数据记录应包括训练数据获取时间、数据来源、数据量统计等内容；
- c) 算法部署应记录部署操作者、时间及相关结果、部署过程脚本、软硬件配置信息等内容。

7.4 算法应用安全

7.4.1 基本安全要求

7.4.1.1 算法公平

App推荐算法服务提供者应满足以下要求：

- a) 为了确保对于数据主体之公平的数据处理，在进行数据抽样时，应确保数据代表性，所有受试者宜有相同的机会在数据中出现，宜将代表性不足的人群添加到训练集中；
- b) 对个人有法律效力或其他重大影响力的算法服务及用户画像不得仅基于个人基因或健康数据等特定数据。

7.4.1.2 算法导向

不得利用算法虚假注册账号、非法交易账号、操纵用户账号或者虚假点赞、评论、转发，不得利用算法屏蔽信息、过度推荐、操纵榜单或者检索结果排序、控制热搜或者精选等干预信息呈现，实施影响网络舆论或者规避监督管理行为。

7.4.1.3 内容呈现

宜建立内容去重、打散干预等策略，并优化检索、排序、选择、推送、展示等规则的透明度和可解释性，避免对用户产生不良影响，预防和减少争议纠纷。

7.4.2 典型场景应用的安全要求

7.4.2.1 信用评级

App推荐算法服务提供者应满足以下要求：

- a) 收集的个人信息应为处理个人征信评价等活动的最小必要个人信息，不应收集相关法律、行政法规规定禁止收集的个人信息；
- b) 向征信机构提供个人不良信息，应当事先告知信息主体本人。依照法律、行政法规规定公开的不良信息除外；
- c) 对达到法律规定保存年限的征信数据应及时进行删除，避免影响公民获得公正的信用评价，防止出现“一处失信，处处受限”的情形。

7.4.2.2 劳动雇佣

提供企业岗位查询与联系等就业机会服务的App应满足以下要求：

- a) 利用算法服务筛选劳动者，不应使用民族、种族、性别、宗教信仰、残疾人、传染病病原体携带者、进城务工的农村劳动者以及乙肝病毒血清学指标等具有歧视性的个人特征，丧失或减少求职者的雇佣机会，或在劳动条件上区别对待，如定向推送的算法决策使得男性比女性看到的高薪招聘广告更多；
- b) 宜合理限制算法服务在人力资源管理领域中的运用目的，除非具有正当“商业事由”，即为订立、履行、解除或终止劳动合同所必需，或按照依法制定的劳动规章制度或依法签订的集体合同实施人力资源管理所必需。

7.4.2.3 网络交易

App推荐算法服务提供者应满足以下要求：

- a) 向消费者销售商品或者提供服务的，应保护消费者公平交易的权利，不得根据消费者的偏好、交易习惯等特征，利用算法对消费者在交易价格等交易条件上实施不合理的差别待遇；
- b) 实施差别待遇的例外情形：
 - 1) 基于消费者实际需求，或符合正当的交易习惯和行业惯例，实行不同交易条件；
 - 2) 基于针对新用户或在合理期限内开展的优惠活动；
 - 3) 基于平台公平、合理、无歧视的规则实施的随机性交易；
 - 4) 基于保障弱势群体利益和安全而做出的限制或差异性对待；
 - 5) 法律法规规定的其他情形；
 - 6) 其他正当性行为。
- c) 宜充分考虑用户体验，科学规划推送频次，不得恶意对不同用户进行差别频次推送。

7.4.2.4 短视频内容

App推荐算法服务提供者提供网络信息内容服务的，应满足以下要求：

- a) 推送短视频内容，应坚持主流价值导向，树立正确的审美观，积极传播正能量，促进推荐算法服务向上向善；
- b) 宜根据正能量内容的类别设置不同的推荐策略，根据级别赋予不同的推荐权重；
- c) 应建立内容过滤的标签体系和研判标准，对违法有害信息进行有效过滤。

8 用户主体权益保护要求

8.1 一般用户主体权利

8.1.1 知情权

App推荐算法服务提供者应满足以下要求：

- a) 应明确告知用户为其提供推荐算法服务的情况，包括对用户画像的使用及基于此处理可能对用户产生的影响等；
- b) 应建立清晰明确、易于用户理解的推荐算法服务公开机制，以适当方式公示推荐算法服务的基本原理、目的意图和主要运行机制。

8.1.2 决定权

App推荐算法服务提供者应满足以下要求：

- a) 通过推荐算法服务方式向个人进行信息推送、商业营销，应同时提供不针对其个人特征的选项，或者向个人提供便捷的拒绝方式；
- b) 宜向用户提供选择或者删除用于推荐算法服务服务的针对其个人特征的用户标签的功能；
- c) 个人有权拒绝个人信息处理者仅通过推荐算法服务的方式作出的对个人权益有重大影响的决定，但下列情形除外：
 - 1) 推荐算法服务对于缔结和履行合同是必要的；
 - 2) 个人明确、单独同意推荐算法服务的运用；
 - 3) 法律、行政法规规定的其他情形。

8.1.3 解释权

通过推荐算法服务作出对个人权益有重大影响的决定，个人要求予以说明的，应当及时说明；说明的范围不限于对个人信息处理规则的解释，还应包括公平性解释（如决策机制公平性以及结果公平性）、影响性解释（如决策如何影响个人以及广泛社会群体、为减轻其所可能带来的负面影响而采取的纠正机制）等。

注：对个人权益有重大影响的决定，指“具备对个人的境况、行为或选择的产生重大影响的潜在可能，在最极端的情况下，可能导致对个人的排除或歧视”的决定，如自动决定个人征信及贷款额度、面试人员的自动化筛选、行政司法决策等。

8.1.4 申诉权

应当建立便捷、有效的投诉处理机制，公开投诉方式等信息，及时受理、处理有关推荐算法服务应用的投诉及反馈处理结果。

8.2 特殊群体权益

8.2.1 概述

本文件中特殊群体包括涉及推荐算法服务的未成年人、老年人、劳动者（提供工作调度服务）群体。除保障一般用户主体权益，还应满足对应群体的要求。

8.2.2 未成年人

主要面向未成年人提供服务的App推荐算法服务提供者，应满足以下要求：

- a) 应通过开发适合未成年人使用的模式、提供适合未成年人特点的服务方式，便利未成年人获取有益身心健康的信息；
- b) 不应向未成年人推送可能引发未成年人模仿不安全行为和违反社会公德行为、诱导未成年人不良嗜好等可能影响未成年人身心健康的信息；
- c) 特定应用场景下，如网络游戏、网络直播、短视频，应设置机制防止未成年人沉迷网络、过度消费，如使用时长提醒、高额消费二次确认等；
- d) 向未满十四周岁的未成年人进行个性化推送时，应优化个性化算法推荐技术，推送适合未成年人年龄的、正向引导其身心健康发展的信息；
- e) 以未成年人为服务对象的在线教育网络产品和服务，不得插入网络游戏链接，不得推送广告等与教学无关的信息。

8.2.3 老年人

面向老年人提供服务的App推荐算法服务提供者应满足以下要求：

- a) 应保障老年人依法享有的权益，充分考虑老年人出行、就医、消费、办事等需求，按照国家有关规定提供智能化适老服务；
- b) 应依法开展涉电信网络诈骗信息的监测、识别和处置，重点避免为他人实施电信网络诈骗提供内容分发、广告推广、引流推广、搜索引擎等网络推广服务。

8.2.4 劳动者

面向劳动者提供服务的App推荐算法服务提供者，应从平台订单分配、报酬构成及支付、工作时间、奖惩等方面保护劳动者的合法权益。不得将最严算法规则作为考核要求，宜建立协商机制，充分听取工会或劳动者代表的意见建议，将结果公示，保障劳动者对涉及切身利益重要事项的知情权。

9 推荐算法安全管理要求

应结合法律法规及国家标准要求，制定推荐算法相关安全管理制度，包括安全运营、安全评估、安全事件应急预案、违规处置等内容。

9.1 安全运营

App推荐算法服务提供者应满足以下要求：

- a) 在日常运营中应按照推荐算法安全管理制度及相应配套规则对算法的输入、设计和实施过程进行运营管理；
- b) 在日常运营中的算法设计环节，应当根据对用户权益可能造成的影响，或者一旦泄露可能对推荐算法系统承载的业务和数据、组织利益及社会公共利益造成的影响，对所设计的推荐算法进行分级评估，根据相应的等级进行分级管理，并定期汇报推荐算法安全风险报告；
- c) 应当定期组织推荐算法相关的安全培训与宣讲，包括但不限于法律法规、政策解读、行业案例等，提高员工的合规意识，提高组织推荐算法的合规水平；
- d) 应当建立检查和监督制度，定期巡检以及对评估为高风险或特定敏感领域的推荐算法不定期抽检。检查内容包括推荐算法的运行情况、数据的处理范围、历史安全事件、推荐算法的定期评估报告、运行日志留存情况、推荐算法安全风险复核等。

9.2 安全评估

App推荐算法服务提供者应满足以下要求：

- a) 利用个人信息开展推荐算法服务，在推荐算法设计和发生重大更新环节，应按照国家有关规定开展安全评估，评估报告和处理情况记录留存期限（不低于3年）。评估的主要内容应包括：
 - 1) 对可预期的处理操作和处理目的进行系统性描述；
 - 2) 对符合数据处理目的的处理行为的必要性和科学性进行评估；
 - 3) 对数据主体享有的权利和自由进行评估；
 - 4) 对可预期的风险所采取的措施进行评估。
- b) 具有舆论属性或者社会动员能力的推荐算法服务应当按照国家有关规定开展安全评估。

注：具有舆论属性或社会动员能力，指提供下列情形的互联网信息服务：开办论坛、博客、微博客、聊天室、通讯群组、公众账号、短视频、网络直播、信息分享、小程序等信息服务或者附设相应功能；开办提供公众舆论表达渠道或者具有发动社会公众从事特定活动能力的其他互联网信息服务。

- c) 应定期开展推荐算法服务的算法输入、设计和实施的自评估，记录并反馈评估结果。

9.3 应急处置

应根据推荐算法服务安全事件对组织系统所承载的业务和数据、组织利益、社会公共利益造成后果的严重程度，对不同等级的安全事件采取相应应急策略。应急预案及处置可参考GB/T 35273中10.1的要求。

9.4 违规追责

在组织的推荐算法服务工作及安全事件应对工作中，发现存在违法违规行为的，应根据安全事件等级，统筹相关职能部门，追究相应部门和人员责任。

10 测试方法

测评主要对App推荐算法服务的个人信息保护、算法模型、算法安全控制、算法典型场景应用、用户主体权益、算法安全管理进行安全评估，评估对象包括App、涉及算法服务的App后台系统、技术措施、

人员组织、过程文档和制度文件等。评估人员和评估机构宜综合采用文档审查、功能核验、人员访谈等测评方式，以测评App推荐算法服务安全性，具体见表1。

表1 测评项目与测评方式对应关系表

要求条款	测评项目	文档审查	功能核验	人员访谈
6	个人信息保护要求	√	√	√
7.2.1	数据处理	—	—	√
7.2.2	用户标签	√	—	—
7.2.3	模型应用	√	√	—
7.3.1	安全审核	√	—	√
7.3.2	干预控制	√	—	√
7.3.3	风险监测	√	—	√
7.3.4	追溯机制	√	—	—
7.4.1.1	算法公平	—	√	√
7.4.1.2	算法导向	—	—	√
7.4.1.3	内容呈现	—	—	√
7.4.2.1	信用评级	—	√	√
7.4.2.2	劳动雇佣	—	√	√
7.4.2.3	网络交易	—	√	√
7.4.2.4	短视频内容	√	√	—
8.1.1	知情权	—	√	—
8.1.2	决定权	—	√	—
8.1.3	解释权	—	√	—
8.1.4	申诉权	—	√	—
8.2.2	未成年人	—	√	√
8.2.3	老年人	—	√	√
8.2.4	劳动者	—	√	√
9.1	安全运营	√	—	√
9.2	安全评估	√	—	√
9.3	应急处置	√	—	√
9.4	违规追责	√	—	√

附 录 A
(资料性)
常见推荐算法应用

常见的推荐算法应用包括但不限于表A.1。

表A.1 常见推荐算法应用

算法技术	应用场景	示例
生成合成类	利用深度学习、虚拟现实等技术，制作文本、图像、音频、视频、虚拟场景等信息，被广泛应用于内容聚合、内容自动化生成及内容合成服务。	写稿机器人、聊天机器人 AI换脸、图像修复 语音合成、自动编曲 视频自动剪辑、合成 游戏、VR、虚拟主播、元宇宙
个性化推送	基于用户画像，以用户标签作为算法的核心变量，针对不同用户需求分发不同内容或提供决策支持，被广泛应用于内容、商品或服务的个性化推送服务。	信息内容自动推送（如新闻、音乐） 广告自动推送（如朋友圈广告） 电商平台的商品推荐
排序精选类	根据指定标准对序列里的数据进行排序，被广泛应用于内容服务提供者的版面页面管理，诸如首屏、热搜、精选、榜单等。	搜索结果排序（如店铺、商品、网页搜索） 热点精选、榜单（如热搜榜、热销榜）
检索过滤类	对大容量数据进行查找，或以特定标准从大容量数据中过滤出某一特定数据，被广泛应用于搜索引擎、内容检索、内容干预等服务。	不良信息过滤 敏感字词识别
调度决策类	根据一定策略分配有限系统资源，在不确定条件下做出最优决策，被广泛应用于资源匹配、工作调度服务。	资源调度（如出行司机调度） 订单分派（如即时配送订单分派） 驾驶决策（如依路况选择最优路径）

参 考 文 献

- [1] 中华人民共和国个人信息保护法（2021年8月20日第十三届全国人民代表大会常务委员会第三十次会议通过）
- [2] 互联网信息服务算法推荐管理规定（2021年11月16日国家互联网信息办公室2021年第20次室务会议审议通过）
- [3] 国务院反垄断委员会关于平台经济领域的反垄断指南（2021年2月7日国务院反垄断委员会印发）
- [4] 具有舆论属性或社会动员能力的互联网信息服务安全评估规定（2018年11月15日国家互联网信息办公室和公安部发布）
- [5] 征信业管理条例（2013年1月21日中华人民共和国国务院令631号公布）
- [6] 中华人民共和国劳动法（2018年12月29日第十三届全国人民代表大会常务委员会第七次会议修正）
- [7] 中华人民共和国就业促进法（2015年4月24日第十二届全国人民代表大会常务委员会第十四次会议）
- [8] 上海市网络交易平台网络营销活动算法应用指引（试行）（2021年11月28日上海市市场监督管理局印发）
- [9] 深圳经济特区数据条例（2021年6月29日深圳市第七届人民代表大会常务委员会第二次会议通过）
- [10] ISO/IEC TR 24027 Information technology — Artificial intelligence (AI) — Bias in AI systems and AI aided decision making
-

电信终端产业协会团体标准

App 推荐算法用户权益保护技术要求及测评规范

T/TAF 138—2022

*

版权所有 侵权必究

电信终端产业协会印发

地址：北京市西城区新街口外大街 28 号

电话：010-82052809

电子版发行网址：www.taf.org.cn