ICS 35. xxx
CCS Lxx

团　　体　　标　　准

T/ISC XXX—XXXX

互联网平台企业推荐算法管理规范

Management Specification for Recommendation Algorithms of Internet Platform Enterprises

在提交反馈意见时，请将您知道的相关专利与支持性文件一并附上。

（征求意见稿）

2025-06-18

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专利的责任。

本文件由中国信息通信研究院提出。

本文件由中国互联网协会归口。

本文件起草单位：中国信息通信研究院、北京快手科技有限公司、阿里巴巴（北京）软件服务有限公司、北京三快在线科技有限公司、同程网络科技股份有限公司、北京小桔科技有限公司

本文件主要起草人：骆曼迪、梁睿琪、王哈达、落红卫、王昕、谷晨、刘艾婧、姜文、郜程程、刘榕、王丹、徐鹏、李媛、薛馨

——

互联网平台企业推荐算法管理规范

1 范围

本文件规定了对互联网平台企业的推荐算法管理要求，包括算法管理原则、算法合规管理组织架构与岗位设置、算法管理制度与流程、算法管理工具配置、算法研发与运行管理、算法测试与评估、安全保障与应急处置、算法研发监督与检查、企业员工合规意识培养的要求。

本文件适用于互联网平台企业规范其推荐算法的设计、研发、测试评估、安全保障、用户权益保护等方面的管理工作。本文件中推荐算法包括生成合成类、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。

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 42888—2023 信息安全技术 机器学习算法安全评估规范

GB/T 43698—2024 网络安全技术 软件供应链安全要求

GB/T 45225—2025 人工智能深度学习算法评估

GB/T 45392—2025 数据安全技术基于个人信息的自动化决策安全要求

GB/T 45574—2025 数据安全技术 敏感个人信息处理安全要求

《互联网信息服务算法推荐管理规定》

3 术语和定义

GB/T 42888—2023、GB/T 45574—2025和《互联网信息服务算法推荐管理规定》界定的以及下列术语和定义适用于本文件。为了便于使用，以下重复列出了GB/T 42888—2023、GB/T 45574—2025和《互联网信息服务算法推荐管理规定》中的某些术语和定义。

3.1

推荐算法 recommendation algorithm

推荐算法是指生成合成类、个性化推送类、排序精选类、检索过滤类、调度决策类算法。

[来源：《互联网信息服务算法推荐管理规定》，有修改]

3.2

生成合成信息 generative synthetic information

生成合成信息是指利用虚拟现实、深度学习等技术对文本、图像、音频、视频、场景模型等进行生成或者编辑所得到的信息。

[来源：GB/T 42888—2023，3.7]

3.3

灰度发布 grayscale release

灰度发布是指在某项产品或应用正式发布前，选择特定人群试用，逐步扩大其试用者数量，以便及时发现和纠正其中的问题。

3.4

敏感个人信息 sensitive personal information

一旦泄露或非法使用，容易导致自然人的人格尊严受到侵害或人身财产安全受到危害的个人信息。

[来源：GB/T 45574—2025，3.2]

注：敏感个人信息包括生物识别、宗教信仰、特定身份、医疗健康、金融账户和行踪轨迹等信息和不满十四周岁未成年人的个人信息。

4 符号和缩略语

下列符号和缩略语适用于本文件。

AI Artificial intelligence（人工智能）

5 推荐算法管理原则

5.1 管理原则

互联网平台企业在推荐算法的规划设计、研发、测试评估、上线运行等阶段应遵循以下原则：

a) 坚持主流价值导向：应识别优质、合规内容并加大推荐力度，限制低俗、劣质及违法内容传播；
b) 健全伦理合规审查：应从科技伦理、用户权益保障及公平性等方面检查算法方案是否合规；
c) 严防信息诱导风险：不应利用算法诱导用户产生高风险行为、过度消费、沉迷内容或引发网络暴力等；
d) 建立公开透明机制：应建立公示机制，公示重要算法、规则的基本原理、目的意图和主要运行机制，且位置显著、公示期合理；
e) 强化特殊群体保障：应保障未成年人、老年人、残障人士等特殊群体的权益，提供适合特殊用户群体特点的服务。

5.2 组织战略要求

互联网平台企业应将推荐算法管理纳入战略规划，明确纲领性要求：

a) 应制定算法治理战略目标，与企业社会责任、用户权益保护及合规要求匹配，确保算法应用符合《中华人民共和国个人信息保护法》《中华人民共和国网络安全法》《中华人民共和国数据安全法》《互联网信息服务算法推荐管理规定》等法律法规；
b) 高层管理者宜定期审核战略实施情况，根据技术发展、政策变化及用户反馈调整方向；
c) 宜将算法治理成效纳入绩效考核体系，与相关部门的工作考核挂钩。

6 算法合规管理组织架构与岗位设置

6.1 组织架构

互联网平台企业应建立多层次的算法合规管理组织架构，应确保从战略决策到具体执行的有效衔接，应设立以下职能部门及制度，或具备同等替代机制，具体要求如下：

a) 应成立算法合规管理委员会，并由企业高层领导担任负责人，主要负责制定战略方向和重大决策；
b) 应成立跨部门协作团队，成员包括但不限于法务、风控、业务、数据管理等部门的团队领导者，负责算法管理方面的议题征集、决策及跨部门协作等事务；
c) 应成立各专业方向实施团队，根据企业所涉及的业务设立信息安全、数据安全、反诈、风控等专业团队，负责具体工作实施如监控不良信息、保护用户数据安全等；
d) 应建立例会制度，定期召开会议，解决需要跨团队协作的算法管理问题，会后形成书面记录，对于特别重要的议题，高层领导应出席决策；
e) 必要时宜针对需要跨部门协作的算法管理事务成立专项工作组，快速响应并推进问题的解决。

6.2 岗位设置

互联网平台企业应明确以下关键岗位及职责或设立其他具有相似职能的岗位：

a) 算法合规总监：统筹算法合规工作，审核方案合规性，向企业高层汇报风险；
b) 数据安全工程师：负责数据采集、存储及使用的安全管控，确保数据合规；
c) 算法测试专员：执行算法测试工作，对算法功能、性能和安全性等进行多维度评估；
d) 用户权益保护专员：监督算法对用户权益的影响；
e) 内容审核专员：人工复核推荐内容，过滤违法违规信息；
f) 应急处置专员：负责算法安全事故的响应与处理。

7 算法管理制度与流程

7.1 核心管理制度

互联网平台企业宜订立清晰完备的算法管理制度，具体要求如下：

a) 算法全生命周期管理制度：涵盖设计、研发、测试、上线、迭代及下线全流程；
b) 数据治理制度：包括数据采集授权、加密存储、权限管理及质量审核；
c) 合规审查制度：明确伦理、法律及社会影响的审查标准与流程；
d) 应急管理制度：规定事故分级、响应流程及责任追究机制；
e) 员工行为规范：明确算法研发与管理中的禁止性行为及合规义务。

7.2 关键流程

互联网平台企业宜规范算法研发和安全事件、用户反馈处理流程，具体流程及示例如下：

a) 算法上线审批流程，如：新算法需经算法测试专员测试→算法合规评估→正式上线运行；
b) 算法迭代更新流程，如：研发团队提交变更申请（说明迭代原因、影响范围及合规评估）→算法测试专员测试通过→灰度发布；
c) 用户反馈处理流程，如：用户反馈受理→分配至相应业务的客服人员进行处理→处理结果向用户同步并留存记录；

d) 算法安全事故处置流程，如：事故发生后立即启动应急响应→封存数据和算法→排查漏洞并修复。

7.3 算法研发与管理文档

互联网平台企业宜结合自身情况制定算法研发规范，应在算法研发和管理的全流程中，撰写并保存必要的文档，具体要求如下：

a) 算法需求说明文档：应包括算法需求提出方、算法功能和性能需求、必要性和可行性说明等；
b) 算法概要设计文档：应包括算法的研发时间、研发团队和负责人、整体架构设计、模块划分、基本原理说明、输入输出说明、功能与性能设计、外部接口设计等；
c) 算法详细设计文档：应包括具体模块设计、数据结构、内部接口、模型架构、详细原理、重要函数、重要参数、训练方法等；
d) 算法研发进度记录：应包括每个阶段的研发进度与规划、功能与性能指标完成情况、风险与问题、里程碑等；
e) 算法运行环境说明：应包括算法运行所需的硬件环境、软件环境、参数配置说明；
f) 算法测试记录：应包括对算法进行功能测试、性能测试、安全性测试的测试方法、测试过程、参数配置、测试结果、发现的问题等；
g) 算法优化与版本说明：对算法进行优化和版本迭代时，应记录版本号、版本发布时间、更新内容、问题修复等；
h) 算法运行维护记录：算法上线后的运行维护时间、运维人员、算法版本号、参数配置、错误日志、运行结果等；
i) 算法运行事故记录与定责处理记录：如果算法上线后发生了事故，应详细记录事故发生时间、造成的影响、事故归因与定责、事故处理结果。

8 算法管理工具配置

互联网平台企业宜配置以下工具支持算法管理：

a) 数据安全工具：数据加密软件、数据访问权限管理工具、数据脱敏工具；
b) 算法测试工具：灰度发布系统、对照实验平台、漏洞扫描工具；
c) 内容审核工具：AI 内容识别工具、人工审核工作台、违规内容检测工具；
d) 监控工具：算法输出实时监控系统、用户反馈分析平台、用户异常行为检测工具；
e) 文档管理工具：算法研发与审核文档库、版本控制软件。

9 算法研发与运行管理

9.1 数据安全与质量保障

互联网平台企业对其算法研发和运行中使用的数据应采取全面的安全和质量管理措施，具体要求如下：

a) 应通过合法合规途径采集数据；
b) 应保护用户敏感个人信息，严格遵守个人信息保护法律法规；
c) 应严格控制数据使用权限，确保数据安全；
d) 应建立严格的数据筛选和清洗流程，保证数据有效性和准确性，避免在算法研发和算法模型训练过程中使用具有偏见和歧视性的数据；

e) 宜尽量丰富数据的多样性，为提高算法和模型的泛化能力、降低算法和模型的偏差奠定数据基础。

9.2 算法安全与质量管理

在算法研发阶段中应注意算法安全并做好质量管理，算法安全性包括算法和模型不被恶意利用或攻击、不产生系统性风险、不侵犯用户权益或存在伦理问题，具体要求如下：

a) 在算法研发阶段应将算法的安全性、鲁棒性等纳入考量，宜减少算法和模型漏洞，降低安全风险，典型漏洞及其安全风险包括：
   1) 数据投毒漏洞：训练数据被篡改导致模型输出偏差；
   2) 对抗性攻击漏洞：恶意输入导致算法误判；
   3) 模型窃取攻击漏洞：攻击者通过对模型的查询访问获取模型结构或参数等信息；
b) 应做好算法发布管理，某版本算法发生事故时便于回退、下线。

9.3 用户自主可控

算法运行过程中应保障用户的知情权和选择权，确保用户对个人兴趣、推荐内容、敏感个人信息使用等自主可控，应对用户提供可选择的服务选项，并及时响应用户所作的选择和修改，具体要求如下：

a) 宜基于用户的浏览历史和个人设置的兴趣标签进行内容推荐，推荐内容宜与用户兴趣相关；
b) 通过算法能力向用户推荐商品和服务时，应保障推荐的公平性和多样性；
c) 通过算法能力向用户展示商品评价或服务评价时，应保障展示顺序及内容的客观性；
d) 当算法服务涉及用户敏感个人信息时，应在用户充分知情的前提下提供服务；
e) 应向用户提供关闭个性化推荐的便捷功能，当用户关闭个性化推荐后，应停止提供相应的个性化推荐服务；
f) 基于用户的兴趣标签向用户进行算法推荐的，应向用户提供兴趣标签选择功能，并对具体的内容应提供不感兴趣选项；
g) 宜向用户提供调节各类型内容推荐强度的功能，并逐步按照用户设置的内容推荐强度完成调整。

9.4 防止信息诱导

互联网平台企业不应利用技术方法和平台内容诱导用户发生高风险或不正当行为，应采取措施控制高风险诱导性信息在互联网平台上的传播，具体包括以下方面：

a) 不应利用算法能力诱导用户产生高风险行为；
b) 不应利用推荐算法能力，诱导用户过度消费、借贷消费，应防止未成年用户非理性消费；
c) 不应通过不实陈述、片面或夸大宣传过往业绩、违规承诺收益或者承担损失等误导性描述或强制捆绑，诱导用户进行网络借贷、投资理财等高风险金融操作，严格遵守互联网金融管理要求；
d) 不应利用推荐算法向用户推送或推荐高度同质化内容来诱导用户沉迷，宜采用同类型内容打散等技术方法保持推荐内容的多样性，当用户未明确设置仅对某些类型内容感兴趣时，宜确保推荐内容类型的丰富性；
e) 不应通过推送或推荐煽动性和激发矛盾的内容诱导用户产生负面情绪、发表攻击性言论或参与网络暴力。

9.5 公开透明

9.5.1 建立健全公示机制

针对直接影响用户使用体验和用户权益的重要算法，互联网平台应建立公示机制，具体要求如下：

a) 公示信息应至少包括算法的基本原理、目的意图和主要运行机制等，比如算法的作用目的、应用范围、规则策略等，应确保公示信息清晰、易于理解。
b) 当平台利用算法向用户个性化推荐优惠促销、抽奖、补贴等活动时，应公示活动参与规则、活动时间、优惠券发放数量和范围、中奖比例、补贴对象和力度等基本情况；
c) 当平台利用算法产生商品排行榜时，应公示榜单排序机制机理，如基本原理、排序依据、主要因素等信息，榜单应基于消费者好评率、销量等客观数据产生，如涉及竞价排序，应有显著标识，防止误导用户的消费选择；
d) 公示位置宜显著，便于用户查看，显著公示位置包括：网站首页、手机应用内弹窗、手机应用内设置页面、手机应用内一级页面信息栏、向用户发送的站内消息等；
e) 当算法发生重大变更，应在完成监管部门备案审批等必要流程后，在 2 个工作日内更新公示信息，公示期不宜过短，以保障用户有足够时间查阅；
f) 平台应向用户说明无法领取或使用优惠券、补贴的真实原因，如领取截止时间、领取要求等，涉及平台商业秘密信息的除外；
g) 应保障公示信息的真实性、全面性和可靠性。

9.5.2 热度榜单

对于社交媒体类和搜索类互联网平台，热度榜单生成算法应遵循以下要求：
a) 应确保热度榜单内容和顺序的客观性、公平性；
b) 宜建立负面内容清单，应基于负面内容清单或其他方式，过滤违法违规内容和其他政策文件规定的不适宜大规模宣传的内容；
c) 对于被举报的内容，经过核实存在问题的，应及时采取措施降低内容热度；
d) 应建立人工巡查制度，对内容热度榜单进行人工审核，确保内容合规。

9.6 特殊用户群体

互联网平台企业提供算法推荐服务应保障用户群体如未成年人、老年人、残障人士的权益，具体要求如下：
a) 应向用户提供未成年人模式选项，宜提供老年模式选项，宜提供视障、听障等特殊用户群体使用模式；
b) 在未成年人模式下，应对产品可能影响未成年人身心健康的功能和使用进行限制，如在线时长，宜建立未成年人专属内容池，为未成年人提供适宜的优质内容；
c) 在老年模式下，宜对字体大小、按钮位置、语音控制等功能和设计做出适老化调整，便于老年用户使用；
d) 不应对特殊群体用户提供歧视性内容和服务。

10 算法测试与评估

10.1 算法研发全流程安全保障

互联网平台企业应在其算法研发的各个阶段采取相应的安全保障措施，具体要求如下：
a) 算法上线前应进行离线测试与评估，可采取黑盒测试、白盒测试和灰盒测试等技术方法对算法作全面的功能、性能和安全性测试；
b) 算法测试应检查实际算法实现与算法设计文档是否一致，算法输出结果是否符合设计文档中的描述；

c) 算法上线后宜先进行灰度发布或对照实验，评估新算法的实际效果，如果发现问题应及时对算法进行调整优化，不可将存在安全隐患的算法直接大规模上线运行；
d) 算法上线后宜在必要环节进行数据埋点和关键日志留存，对重要的运行数据进行持续的监测。

10.2 算法评估维度及量化指标

算法评估维度及量化指标应满足以下要求：
a) 应评估算法输出、推荐或推送的内容是否健康、积极，是否存在不良信息；
b) 应评估算法是否侵犯用户的隐私权、公平交易权和其他合法权益；
c) 应评估算法是否存在被滥用风险，如被用于生成虚假信息、诈骗信息等；
d) 应评估算法是否存在技术漏洞，导致在特定输入下产生错误或有害的输出，影响算法的可靠性和安全性；
e) 宜建立客观、全面、多样的量化指标以衡量算法的实际效用，对于推荐类算法，其量化指标宜包括但不限于：
   1) 内容覆盖率：衡量推荐系统覆盖的内容广度；
   2) 推荐准确率：评估推荐内容与用户兴趣的匹配度；
   3) 用户满意度：宜通过技术方法或问卷调查来获取用户的直接反馈。

10.3 用户采样

算法上线运行后，宜对用户群体进行采样调查，以评估算法实际产生的效用，具体要求如下：
a) 采样调查宜公平客观的选取参与调查的用户，并采取科学的统计方法保障调查的公平性和有效性；
b) 宜采样调查用户是否受到不公平的差别待遇，如收到的优惠券、参与平台活动资格等；
c) 宜采样调查用户使用体验是否受到显著影响，如是否收到推送消息滋扰、广告弹窗滋扰等；
d) 如果发现算法对用户权益和使用体验产生负面影响，应及时采取措施进行补救。

11 安全保障与应急处置

11.1 算法安全保障制度

应建立完善的算法安全保障制度体系，具体要求如下：
a) 应建立数据安全管理规范、算法需求与设计审查规范、算法测试与评估规范、算法安全事故处理规范等制度规范，保障平台算法安全性和合规性；
b) 应建立内部安全规范的制定、更新和废除机制，保障安全制度的有效性和可实施性；
c) 应设立用户反馈建议和申诉的渠道。

11.2 算法安全事故预防机制

应综合采取技术监控方法和人工巡查机制预防算法安全事故的发生，具体要求如下：
a) 宜采用异常模式或行为识别、规则引擎与动态阈值等技术手段监控平台指标，防止信息屏蔽、过度推荐、热度榜单操纵等问题出现；
b) 应结合技术监控与人工审核的结果进行综合判断，更准确地识别潜在的安全威胁。

11.3 风险提示

应对用户使用过程中存在潜在风险的内容和事项进行风险提示，具体要求如下：

a) 对AI生成合成信息应依法标注AI显式标识；
b) 在夜间时段或未成年人模式下宜设置防沉迷提示；
c) 对股票讲解和其他投资理财类内容宜设置风险提示；
d) 对涉宗教信仰类内容宜设置标识或提示；
e) 在未成年人模式下，宜对存在模仿行为风险的内容设置危险提示。

11.4 事故应急处置

应建立算法安全事故应急响应机制，以保障平台的平稳运行并减少算法安全事故对用户的影响，具体要求如下：

a) 应根据对国家安全、社会秩序、经济建设和公众利益的影响程度建立明确的算法安全事故分级分类标准，针对不同类别和性质的事故应急处置工作，宜明确各部门的责任分工并制定详细的应对措施，如在事故处理过程中发现问题较为严重，应升级处理；
b) 宜建立事故警报机制，在事故发生后及时通知相关部门负责人；
c) 宜在事故处理后排查并确定事故原因和影响范围，采取修复措施以防止事故复发。

12 算法研发监督与检查

12.1 内部监督

互联网平台企业宜设立健全的内部监督机制，具体要求如下：

a) 宜开展合规检查，宜重点核查公示机制、用户权益保护及数据安全措施；
b) 宜审核算法研发和运行全流程，并形成报告提交算法管理委员会；
c) 宜设立内部算法合规举报渠道，打击违规行为。

12.2 外部检查响应

互联网平台企业应配合有关监管部门的检查工作，具体要求如下：

a) 应配合监管部门检查，并按要求提供算法研发相关的说明和备案材料；
b) 宜接受第三方机构评估并进行改进；
c) 如出现重大问题，宜公开检查结果与整改措施，接受社会监督。

13 企业员工合规意识培养

应向企业内部涉及算法规划、设计、研发、运维、管理等工作的员工提供算法合规方面的教育培训和学习材料，具体要求如下：

a) 宜组织关于用户敏感个人信息保护、数据安全、知识产权、用户权益保护等方面法律法规和政策的学习培训，确保员工了解算法合规要求；
b) 宜在企业内部公示通告互联网行业算法违规案例，通过具体的案例剖析，使员工理解不合规行为可能带来的法律后果；
c) 宜建立企业内部学习平台，使员工能够便捷地获取算法合规方面的学习材料；
d) 宜在算法研发和管理的关键岗位新员工入职时提供必要的算法合规意识培训。