ICS 33.030
CCS M 21

团　体　标　准

T/TAF 227—2024

推荐算法干预控制实施指南

Implementation guidelines for recommended algorithmic intervention and control

2024-05-13 发布　　2024-05-13 实施

电信终端产业协会　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、北京抖音信息服务有限公司、北京快手科技有限公司、蚂蚁科技集团股份有限公司、阿里巴巴（中国）有限公司、OPPO广东移动通信有限公司、北京微梦创科网络技术有限公司、上海携程商务有限公司等。

本文件主要起草人：武琳娜、王艳红、王淞鹤、田申、李昳婧、杜蕾、刘笑岑、杨骁涵、谷元坤、吴少卿、落红卫、谷晨、石玉珍、黄天宁、陈志军、任资政、王天、胡立平、高任宇、刘扬等。

引　　言

《个人信息保护法》第二十四条中规定，“个人信息处理者利用个人信息进行自动化决策，应当保证决策的透明度和结果的公平公正”。《互联网信息服务算法推荐管理规定》第十一、十二条以及《网络信息内容生态治理规定》等法律法规进一步细化要求，算法推荐服务提供者应当建立完善人工干预和用户自主选择机制。在推荐算法管理领域，我国已在法律与部门规章等规范性法律文件补足监管空白的背景下，依然面领着标准不清、实践不明等问题，为本实施指南的订立提供了必要性基础。本指南的目的系为人工干预和用户自主选择机制的建立和完善提供可参考的标准。

推荐算法干预控制实施指南

1 范围

本文件给出了推荐算法干预控制的原则与实施方式，包括具有可操作性的人工干预机制，以及用户自主选择机制和透明度，并结合行业应用场景给出可参考的示例。

本文件适用于算法推荐服务提供者参考建立完善人工干预和用户自主选择机制，也可为第三方测评机构对推荐算法干预控制过程的评估提供参考。

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 25069—2010 信息安全技术 术语

GB/T 35273—2020 信息安全技术 个人信息安全规范

T/TAF 138—2022 APP推荐算法用户权益保护技术要求及测评规范

3 术语和定义

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。

3.2

干预控制 intervention control

推荐系统的服务提供者通过人工等方式介入推荐算法流程，调整相应规则和配置的过程，或直接调整推荐结果的过程。

3.3

内容召回 recall

指按照相对简单的逻辑从推荐候选集中快速选择尽可能多的正确结果，并将结果返回给模型进行排序。

3.4

预估目标 objective

推荐系统预期达到的效果指标，如点击率、有效播放率、播放时长、点赞率、关注率等。

3.5

内容消重　deduplication

推荐系统根据用户的历史、特定规则去除候选内容中重复内容。

3.6

融合公式　value tree

推荐系统里面使用的一个技术，用来计算模型的融合分数。

3.7

助推规则　boost

在预估服务中，通过调整预估评分的权重影响结果的排序。

3.8

打散　shattering

对相同内容标签、功能模式、话题类型的内容进行打散，以避免推荐内容的单调性，防止同类内容过于密集。

3.9

强插　forced insertion

在最终推荐混排流的指定位置展示特定内容。

3.10

混排　Re-ranking

不同的推荐内容在最终向用户进行展示前进行整合、分配的策略。

4　基本原则

推荐算法干预控制需考虑以下基本原则：

a) 主流价值导向：坚持算法推荐的主流价值导向，针对优质、正向内容和负面、消极内容适配不同的算法推荐干预措施，以实现积极传播正能量；

b) 准确性与多样性：算法干预措施的选取宜有助于算法准确率和召回率的实现，并进一步拓展内容的丰富度；

c) 公平性：算法干预措施宜有助于实现社会的公平、公正及对特殊群体的重点保护，在复杂的算法推荐生态系统中寻求各方利益和诉求的平衡；

d) 个人权益保障：在算法系统设置干预措施之外，还可配备信息主体对于算法推荐结果的控制措施从而实现对于个人权益的保障。

5　具体措施与实施要求

5.1　算法干预控制机制

推荐系统一般由内容召回、过滤消重以及预估排序等多个相互联结的自动化环节构成，算法推荐服务提供者可在上述环节中通过人工或其他规则方式介入系统，影响最终的推荐结果。推荐系统的主要环节以及重要的干预控制措施见图1。

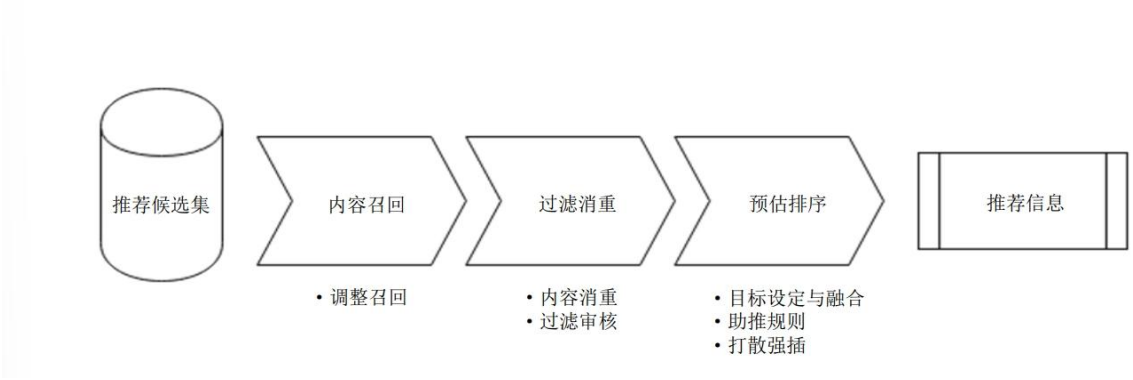

图1　推荐系统的主要环节以及重要干预控制措施

5.1.1　内容召回

在内容召回阶段，推荐算法按照特定的规则和逻辑从全量的信息候选集中召回信息进入预估、排序阶段。召回阶段宜满足如下要求。

5.1.1.1　召回内容标签

召回内容标签的设置合法合规，不得将违法和不良信息作为内容标签，使用个人的偏好特征作为召回内容标签时，不得侵犯个人的权益。

5.1.1.2　召回策略

内容召回策略需考虑以下要点：

a)　信息内容安全：不设置可能增加违法、不良信息召回概率的策略；

b)　用户权益保障：不设置诱导用户沉迷、过度消费的召回策略，不设置可能侵害未成年人、老年人权益的召回策略，导致对个人在交易价格等交易条件上实行不合理的差别待遇；

c)　市场竞争秩序：不利用召回策略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；

d)　公平性：保障算法公平性，使同等质量的合法信息及内容得到公平的召回。

5.1.2　过滤消重

推荐候选集的内容经过召回后，推荐系统宜对内容进行过滤和消重，使违法违规、不良信息和重复内容不被推荐。过滤、消重是对推荐内容进行的主动筛选，宜满足以下要求：

a)　不断优化过滤审核模型和相关技术，提升过滤审核的准确度和有效性，及时发现、识别、过滤和处理违法、不良信息，防止信息扩散；

b)　建立健全用于识别违法和不良信息的特征库，完善入库标准、规则和程序；

c)　合理利用内容消重策略，避免同质化内容的过度推荐，增强内容推荐的多样性；

d)　不宜利用内容消重策略，影响信息内容的公平分发，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。

5.1.3　预估排序

5.1.3.1　概述

在预估排序阶段，推荐算法通过机器学习等模型预测待推荐信息的目标评分，并按照评分对待推荐信息进行排序，按照顺序规则输出推荐信息，预估排序阶段需考虑以下要点。

5.1.3.2 目标设定与融合

预估排序的目标设定与融合需考虑以下要点：

a) 推荐系统设定的预估目标合法合规，不设定可能导致违法、不良信息传播，可能导致过度推荐、不合理的差别待遇以及其他违法违规、违背伦理或公平性的目标；
b) 推荐系统的多个目标可通过计算公式融合最终预估分，融合公式的目标参数和权重配置应保证推荐结果的公平性，不宜将可能诱导用户沉迷、过度消费的目标配置过高的权重。

5.1.3.3 助推规则

预估排序的助推规则需考虑以下因素：

a) 信息内容安全：不设置对违法、不良信息进行推荐的助推规则；
b) 用户权益保障：不设置诱导用户沉迷、过度消费的助推规则，不宜设置可能侵害未成年人、老年人权益的助推规则，导致对个人在交易价格等交易条件上实行不合理的差别待遇；
c) 市场竞争秩序：不利用助推规则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，宜兼顾算法的公平性，使同等质量的合法信息及内容得到公平的分发；
d) 公平性及多样性：保障算法公平性，使同等质量的合法信息及内容得到公平的分发。减少可能导致同质化内容过度推荐的助推策略，宜设置增加内容多样性的助推策略。

5.1.3.4 打散强插

预估排序的打散强插需考虑以下要点：

a) 合理设置打散、强插策略和标准，减少同质化内容的过度推荐，增强内容多样性，保障信息的公平分发；
b) 推荐的内容进行混排时，合理分配不同推荐系统推荐内容，平衡自然推荐信息和营销信息的比例；
c) 不强插违法、不良及侵权信息，避免可能强插诱导用户沉迷、过度消费或侵犯未成年人、老年人权益的策略。

5.2 用户自主选择机制

5.2.1 概述

算法推荐服务提供者宜保障用户对推荐算法及其处理的数据的管理控制权以及自主选择权，用户可以选择退出、关闭推荐算法决策，可以调整、修改部分算法的设置，影响算法的决策结果。

5.2.2 关闭算法推荐服务

算法推荐服务提供者向用户提供不针对其个人特征的选项，或向用户提供便捷的关闭个性化推荐算法的选项。例如，App通过应用内的一键开关提供便捷的关闭功能。

用户选择关闭个性化推荐算法服务的，算法推荐服务提供者需立即停止提供相关服务，但仍然可以非个性化的方式进行内容推荐。如随机选取内容、热门内容或根据统一的排序进行推荐。

5.2.3 用户标签、特征管理

算法推荐服务提供者对于用户标签、特征管理需考虑以下要点：

a) 可向用户提供选择或者删除用于算法推荐服务的针对其个人特征的用户标签的功能，例如用户自主选择或删除偏好商品或内容标签；

b) 不依赖用户标签进行内容推荐的算法，宜向用户提供撤回部分或全部个人信息用于算法推荐的功能，如为用户提供撤回部分或全部个人信息用于算法推荐的开关，或提供不收集部分个人信息的无痕模式。

5.2.4 内容类型管理

算法推荐服务提供者可为用户提供针对某类内容进行反馈或投诉的渠道，以便用户实现对某类内容的屏蔽或减少。宜为用户提供探索更多类型内容的功能，增强内容多样性。如为用户提供自主选择兴趣内容的功能，帮助用户拓展兴趣的功能等。

5.3 透明度要求

5.3.1 系统解释

算法推荐服务提供者可通过产品的个人信息处理规则等方式，以清晰易懂的语言向个人详细披露推荐算法中干预控制机制以及用户自主选择机制的具体情况。

5.3.2 个案解释

对于个人针对单个具体推荐内容进行反馈、投诉，算法推荐服务提供者宜及时回复、解释。宜通过产品功能页面等途径向用户即时告知和披露某单个具体内容被推荐的机制、原因或反馈方式。

个性化推荐算法对个人权益有重大影响的决定，个人有权要求算法推荐服务提供者予以说明。

附 录 A
（资料性）
典型场景的特别要求

推荐算法的应用场景广泛，涉及图文、视频、商品、广告等不同对象，覆盖个性化推荐、信息检索、精选排序等不同类型。典型场景的特别实践要求见表A.1。

表A.1 典型场景实践要求

典型场景	算法干预控制机制	用户自主选择机制
图文、视频内容推荐、搜索	不宜导致违法、不良内容的推荐、传播。 增强内容的多样性，避免同质化内容过度推荐。	宜提供关闭算法推荐服务的选项。 宜提供用户标签或特征管理的功能。 宜提供内容管理、反馈的功能。
商品推荐、搜索	不宜导致大数据杀熟等差别待遇。 不宜诱导用户沉迷、过度消费。	宜提供关闭算法推荐服务的选项。 宜提供用户标签或特征管理的功能。
广告推荐	不宜导致违法、不良广告内容的推荐、分发。 不宜诱导过度消费。 不宜导致差别待遇。	宜提供关闭算法推荐服务的选项。 宜提供用户标签或特征管理的功能。
热榜推荐	不宜操纵榜单、控制热搜或者精选等干预信息呈现。 不宜实施影响网络舆论或者规避监督管理行为。	宜提供用户反馈的机制。 宜对广告、推广内容进行特别标识。

电信终端产业协会团体标准

推荐算法干预控制实施指南

T/TAF 227—2024

*

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