ICS 35．240
CCS L67

YD

中 华 人 民 共 和 国 通 信 行 业 标 准

XX XXXXX—XXXX

基于人工智能的移动应用视频推荐服务技术要求

Technical requirements for AI-based APP video recommendation service

（报批稿）

XXXX－XX－XX 发布　　XXXX－XX－XX 实施

中华人民共和国工业和信息化部　　发 布

目　　次

前　　言 II
1 范围 4
2 规范性引用文件 4
3 术语和定义 4
4 缩略语 4
5 技术框架 5
6 视频推荐服务人工智能通用要求 5
6.1 数据 5
6.2 模型 6
7 功能要求 6
7.1 智能分析模块 6
7.1.1 视频内容理解模块 6
7.1.2 用户分析模块 7
7.2 搜索查询处理模块 8
7.3 视频推荐决策模块 8
7.3.1 召回模块 8
7.3.2 排序模块 9
7.3.3 决策生成 9
7.4 视频内容呈现模块 10
7.5 评估反馈模块 10
8 业务流程 10
8.1 视频搜索推荐 10
8.2 视频展示推荐 11
参考文献 12

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。

本文件由中国通信标准化协会提出并归口。

本文件起草单位：北京快手科技有限公司、中国信息通信研究院、北京抖音信息服务有限公司、蚂蚁科技集团股份有限公司、北京邮电大学、中国移动通信集团有限公司。

本文件主要起草人：落红卫、谷晨、傅山、杨帆、李宣平、王刚、陈赢、冷德维、李永昌、吴少卿、林冠辰、张勇、金矾。

基于人工智能的移动应用视频推荐服务技术要求

1 范围

本文件规定了基于人工智能的移动应用视频推荐服务的技术框架、人工智能通用要求、功能要求和业务流程。

本文件适用于对基于人工智能的移动应用视频推荐服务进行设计、研发和测评，也可为监管部门、第三方评估机构对基于人工智能的视频推荐服务活动进行监督、管理、评估提供参考。

2 规范性引用文件

下列文件对于本标准的应用是必不可少的。凡是注日期的引用文件，仅注日期的版本适用于本标准。凡是不注日期的引用文件，其最新版本（包括所有的修改单）适用于本标准。

GB/T 35273—2020　　信息安全技术 个人信息安全规范

3 术语和定义

下列术语和定义适用于本文件。

3.1

基于人工智能的移动应用视频推荐服务 AI- based APP video recommendation service

采用人工智能技术，分析、评估、预测用户的特征，分析视频服务平台的视频内容，选取与用户匹配的视频，并向用户展示相关视频的APP服务。

3.2

嵌入向量 embedding vector

将对象转换为多维向量，用以表示对象的特征。

3.3

索引池 index pool

与业务场景匹配的，可推荐视频的全集。

4 缩略语

下列缩略语适用于本文件。

APP	移动应用	Mobile Application
OCR	光学字符识别	Optical Character Recognition
IP	知识产权	Intellectual Property
MECE	相互独立，完全穷尽	Mutually Exclusive,Collectively Exhaustive

5 技术框架

基于人工智能的移动应用视频推荐服务系统的技术框架如图 1 所示

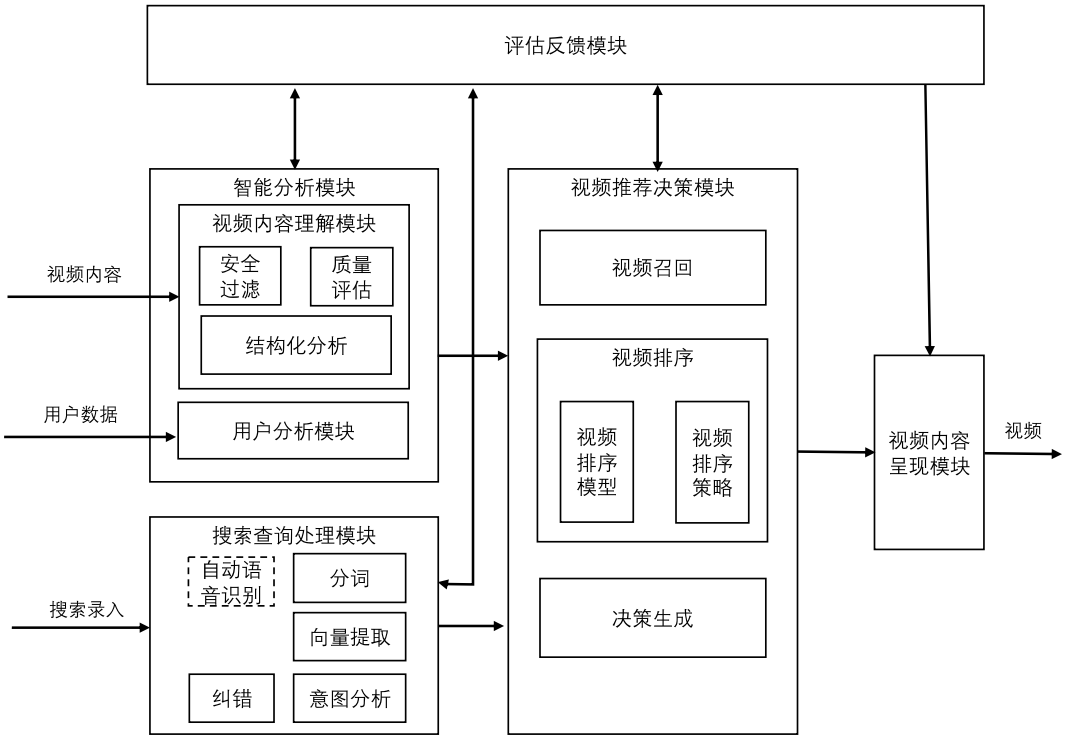

图1　技术框架

人工智能在视频推荐服务中可应用于：

a)　智能分析模块；

b)　搜索查询处理模块；

c)　视频推荐决策模块；

d)　视频内容呈现模块；

e)　评估反馈模块。

视频内容理解模块、搜索查询处理模块、视频召回、视频排序应部署于平台侧实现。

用户分析模块、决策生成、视频内容呈现模块可部署于APP侧，也可部署于平台侧，也可在APP侧和平台侧同时部署实现。

6　视频推荐服务人工智能通用要求

6.1　数据

对人工智能算法处理的数据，应满足如下要求：

a)　应对训练数据进行预处理，包括识别并删除恶意样本、识别并修复或过滤被污染数据；

b)　宜对训练数据集采取必要的保护措施，确保数据的保密性、完整性；

c)　当采用对比学习时，应减少负样本训练数据的随机性，如将目标用户没有正向反馈但其他用户有正向反馈的视频作为单独的负样本训练数据集；

d)　宜构建不同于训练数据集的标准测试数据集，模型上线前应通过标准测试数据集的测试；

e)　应构建合理的训练数据集，包括但不限于数据集的规模、均衡性、准确性等。

6.2　模型

应用于视频推荐服务的人工智能模型应满足以下要求：

a) 应定期针对模型机制机理、应用结果进行审核、评估，并对模型训练和模型推理进行动态改进和调整；
b) 不得设置诱导用户沉迷、过度消费等的人工智能模型；
c) 对风险点进行监测；
d) 宜支持主流算法和模型，包括线性回归、逻辑回归、线性判断分析、朴素贝叶斯、随机森林、多目标、深度神经网络、强化学习等；
e) 宜支持模型训练优化，如模型组合、参数调整、实验调优等；
f) 模型宜具有可解释性。

7 功能要求

7.1 智能分析模块

7.1.1 视频内容理解模块

7.1.1.1 通用

基于人工智能的视频内容理解应满足以下要求：

a) 应生成针对视频内容的标签，并支持标签的更新；
b) 宜支持生成视频内容的嵌入向量；
c) 应支持多模态内容理解，包括但不限于视频、文本、图像、作者、语音、音乐、虚拟场景等；
d) 宜支持前端处理技术；
e) 应支持过滤机制的调整。

7.1.1.2 安全过滤

基于人工智能的安全过滤应满足以下要求：

a) 应能识别、过滤违法内容以及不良内容；
b) 应支持采取措施阻断视频的传播。

7.1.1.3 质量评估

基于人工智能的质量评估应满足以下要求：

a) 应支持对视频的质量评估，至少识别出劣质、普通、优质。应根据质量评估结果采取相应的处理措施，如对劣质视频进行限流等；

注：劣质视频可包括画面感官不适的视频、营销感强烈、格调低下、内容荒诞的视频等。

b) 应支持相似视频的识别，满足以下要求：
    1) 采用的相似度判断算法应具有一定鲁棒性，包括能识别修改分辨率、裁剪、混剪、加片头片尾、修改视频背景、特效处理、滤镜处理；
    2) 训练数据应进行颜色（亮度、对比度、饱和度等）、空间（翻转、旋转、平移等）、时间（慢放、快放、倒放等）等多维度数据增强；
    3) 宜结合全图图像特征和局部图像特征进行相似度比较；
    4) 注：局部图像特征通常针对图像帧的核心运动区域；
    5) 宜支持不同视频的目标帧序列之间的联合相似度比较；
    6) 宜支持画中画区域的定位，并针对画中画区域识别相似视频；
    7) 宜支持采用图理论提升相似视频识别的鲁棒性。

7.1.1.4 结构化分析

基于人工智能的结构化分析应满足以下要求：

a) 应支持多级类目标签，宜遵循 MECE 原则，同级类目标签之间相互独立，同级类目标签之和无遗漏的覆盖上一级的相应类目标签；

注：如运动/球类运动/篮球

b) 应支持多维特征标签，如内容主题标签、内容元素标签、风格标签、内容形式标签、展示形式标签、IP 标签、商品标签、兴趣点标签、实效性标签等。不同维度的特征标签可交叉重叠，可存在同义关系、相关关系、上下位关系等；

注：内容元素标签如人物/场景等；风格标签如小清新/中国风等；内容形式标签如探店/开箱等；展示形式标签如动画/街访等；IP标签如影视IP三国演义等；商品标签如连衣裙等；兴趣点标签如颐和园玉带桥；实效性标签如强/弱/无等。

c) 应支持重复视频的识别，并进行相应的处理；

d) 宜支持对知识视频的识别，通过知识图谱（多模态实体链接、层次聚类、知识点匹配）等技术，构建体系化的词条上下位关系、知识点关系，形成结构化知识视频并进行相应挂载；

e) 宜支持百万级别训练数据；

f) 模型宜支持多任务；

g) 宜支持百亿规模参数内容统一建模。

7.1.2 用户分析模块

7.1.2.1 用户权益保护

基于人工智能进行用户分析时，应满足以下要求：

a) 对用户的个人信息保护应满足 GB/T 35273-2020 的要求；

b) 应告知向用户提供视频推荐服务的场景、相应所需处理的个人信息、处理方式和目的；

c) 处理未成年人个人信息，应公开专门处理规则，明示未成年人个人信息处理的目的、方式和范围；

d) 处理儿童个人信息，应征得儿童的父母或其他监护人同意；

e) 应以显著方式告知用户拒绝提供视频推荐服务所必需个人信息时的后果；

f) 宜使用基于用户群体特征标签的简介画像信息进行视频推荐服务；

g) 应以显著便捷的方式向用户提供偏好选择的功能，宜提供给用户不同级别的推荐强度选择；

h) 应以显著方式向用户提供个人信息获取权限的打开或关闭方式；

i) 应支持用户关闭个性化视频推荐服务。

7.1.2.2 功能要求

基于人工智能的用户分析应满足以下要求：

a) 支持用户选择、删除标签；

b) 生成针对用户个人特征的标签；

c) 用户标签不应包含有对民族、种族、宗教、残疾、疾病、人群行为等歧视的内容；

d) 注：人群行为如低消费水平用户等。

e) 用户标签不应包含负面含义的内容；

f) 应对训练数据中涉及用户个人敏感信息的数据进行脱敏处理；

g) 用户标签涉及敏感个人信息的，应建立脱敏处理机制。

7.2 搜索查询处理模块

搜索查询处理模块应支持以下功能：

a) 应支持文本、语音等形式的搜索查询；

b) 对于语音搜索查询，通过自动语音识别转换为文本；应保持搜索查询的一致性，即语音搜索查询的结果，与直接以同样文本进行搜索查询的结果保持一致；

c) 应支持搜索查询的纠错；

d) 应支持搜索查询的意图分析，包括以用户为搜索查询目标的意图和以视频内容为搜索查询目标的意图。

7.3 视频推荐决策模块

7.3.1 召回模块

7.3.1.1 视频展示推荐

对于展示推荐的视频推荐召回，应满足以下要求：

a) 从索引池海量视频中检索筛选出匹配用户兴趣的视频；

b) 应支持多因素召回，包括视频内容、视频类别、用户、作者、用户与作者的关系等；

注：如向用户推荐展示其粉丝用户的视频。

c) 应支持主流的协同过滤方案，包括基于物品（视频）的协同、基于用户的协同、用户到物品（视频）的协同等；

注：实践中通常以用户到物品（视频）的协同为主要召回通路。

d) 宜支持基于用户与视频交互行为生成嵌入向量；

注：实践中通常将显式正反馈（如点赞、转发等）交互行为设置更高的权重，将隐式正反馈（如观看）交互行为设置较低的权重。

e) 应支持多路召回，可灵活配置各路召回的召回规模，如直接设置个数，或是阈值设置等

f) 宜支持在线实时的召回搜索；

g) 宜支持实时训练、实时更新；

h) 应提供业务规模匹配的召回处理能力。

7.3.1.2 视频搜索推荐

对于搜索推荐的视频推荐召回，应满足以下要求：

a) 从索引池海量视频中检索筛选出匹配用户输入的搜索查询的视频；

b) 应支持主流的召回方式，包括倒排索引、向量检索、协同过滤等；

c) 应支持多路召回，应具备配置各路召回规模上限的能力；

注：实践中通常以倒排索引为主要召回通路，协同过滤召回规模较小。

d) 在倒排索引中，视频特征应包括视频的自包含文本（如封面文本、标签文本、描述文本等）、OCR 识别文本、内容理解标签等；

e) 应支持基于用户历史行为序列的个性化视频搜索推荐；

f) 宜支持在线实时的召回搜索；

g) 宜支持实时训练、实时更新。

7.3.2 排序模块

7.3.2.1 视频排序模型

基于人工智能的视频排序模型，应根据业务特点，采用满足需求的排序模型：

a) 当用于粗排时，为了确保资源能够支撑快速从召回视频池中进行筛选，宜采用复杂度相对较低的模型，例如基于物品（视频）的协同相似度分析、双塔模型等。模型使用特征相对较少，耗时较短，吞吐量较高；
b) 当用于精排时，为了确保能够从粗排视频池中进行精准筛选，宜采用复杂度相对较高的模型。模型使用特征相对较多，结果精确度较高；
c) 宜支持百亿规模的用户行为序列样本特征规模；
d) 宜支持实时训练、实时更新；
e) 对于视频搜索，模型特征应包括视频、视频自包含的文本、视频内容理解标签、视频内容理解嵌入向量、视频自包含图片信息（如封面）、以及搜索查询向量等；
f) 对于视频搜索，应支持不同类目的视频按类目相关性排序，宜支持不同类目视频推荐数量的均衡。

7.3.2.2 视频排序策略

基于人工智能的视频排序，应满足以下要求：
a) 对于视频展示推荐业务，应支持多因素排序，如内容理解、作者表征、关注关系、用户行为、业务场景等；
b) 应支持多目标排序。应包括时长类（如播放时长、完成度、有效播放率、完播率等）、互动类（如点赞率、关注率、分享率、评论率等）、营收类（如打赏率、下单率、商品交易总额）等；对于视频搜索的业务，还应包括搜索查询相关性、权威性、时效性等；
c) 应支持多目标的动态优化调整；
d) 应支持个性化排序策略；
e) 应支持多目标的融合，采取归一化等方法消除不同目标量级差异、分布差异等影响。
f) 宜支持内容风控策略。

7.3.3 决策生成

决策生成模块根据不同的服务类型进行决策，应满足以下要求。
a) 应支持人工干预，包括但不限于基于业务规则和价值观的扶持、排名排序干预、内容打压等；
b) 应支持识别相似视频内容，减少相似内容的重复推送；
c) 对于非搜索推荐业务，应支持多维度的打散干预。如减少重复类型内容的推送，确保视频类目的多样性；控制同一作者内容推荐的频率；控制类似特征标签的视频内容的推送等；
d) 对于非搜索推荐业务，宜支持适量引入随机内容；
e) 对于非搜索推荐业务，宜支持根据网络环境进行推荐视频的缓存，并在网络异常时对缓存视频进行排序，按序向用户推送缓存的视频；播放过的视频宜从缓存中清除；宜优先推送上一次网络异常后缓存的视频；
f) 不应设置诱导用户沉迷、过度消费的人工智能模型；
g) 不应恶意屏蔽信息、不应恶意操纵榜单以及搜索结果等；
h) 应支持推荐视频的重排；
i) 宜支持端侧重排调整；
j) 宜支持生成辅助展示信息，如热度、推荐理由等；
k) 宜支持针对未成年人模式、适老模式等提供符合该人群特点的视频内容；
l) 应支持重排后原生视频与其他领域场景（如广告营销服务、电商推荐服务等）的混排
   1) 原生视频可包括自然流量短视频、冷启推荐、关系链推荐等；

2) 混排应考虑业务约束规则、用户标签、用户体验等；
3) 混排宜考虑原生视频与其他服务的关系，支持过滤及重计算；
4) 应设置其他服务与原生视频的最大比值；

m) 应支持的视频推荐服务包括但不限于：
1) 视频内容的个性化推送；
2) 视频内容的排序精选；
3) 视频内容的检索过滤。

7.4 视频内容呈现模块

基于人工智能的视频内容呈现应满足以下要求：

a) 应显著标示个性化呈现的内容；
b) 应提供退出、关闭个性化视频内容呈现的方式；
c) 应根据用户设置的时间管理或时间提醒，在设置的条件触发时向用户推送提醒内容。

7.5 评估反馈模块

应设计合理的评价体系，观测评价指标，反馈改进人工智能模型。

a) 应在模型上线部署前进行评估，并在上线后定期进行评估反馈，必要时触发模型的优化；
b) 宜支持实时训练和优化；
c) 对于内容理解模块的类目标签，应根据准确率和覆盖率进行反馈和优化；
d) 对于内容理解模块的特征标签，应根据准确率、覆盖率、丰富度进行反馈和优化；
e) 对于内容理解模块的特征标签，宜对不同视频类型设置不同的反馈优化指标；
f) 对于视频推荐决策，应支持综合考虑正向、反向反馈进行优化；应综合考虑显式反馈、隐式反馈进行优化；
g) 对于视频搜索推荐，应支持客观指标进行准确性的评估反馈；
h) 对于视频推荐结果，宜从效率指标，用户体验指标，内容新颖度指标等角度进行推荐结果的评估。

8 业务流程

8.1 视频搜索推荐

视频搜索推荐业务由用户的搜索查询输入触发，结合对用户偏好的理解，向用户返回搜索结果。流程包括搜索查询理解、召回、粗精排、重排、混排、结果呈现，如图2所示。

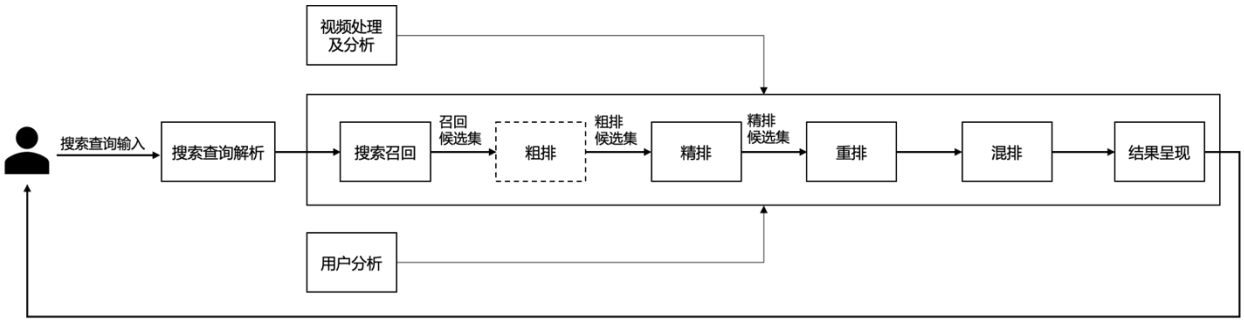

图2 视频搜索推荐流程图

其中：

a) 搜索查询解析：分析用户输入的文本形式或语音形式的搜索查询，由搜索查询处理模块完成；
b) 搜索召回：基于搜索查询的分析结果，在搜索推荐场景下的索引池中，利用人工智能技术检索出符合搜索条件的、满足用户偏好的视频集合，生成召回候选集。由视频召回模块完成；
c) 粗排、精排、重排：在召回候选集中，基于视频内容的理解和用户分析，完成针对搜索查询的推荐视频的确定和排序。由视频推荐决策模块中的视频排序、决策生成共同完成：
   1) 粗排使用的视频排序策略目标相对较少。粗排的性能效率相对要求较高；
   2) 精排使用的视频排序策略目标相对较多。精排的结果准确性要求较高；
d) 混排：将经过粗排、精排、重排的原生视频与其他领域视频进行混合编排。由视频推荐决策模块中的决策生成完成；
e) 结果呈现：将最终的视频推荐结果返回呈现给用户。由视频内容呈现模块完成。

8.2 视频展示推荐

视频展示推荐业务结合对用户偏好的理解，向用户展示推荐的视频。流程包括召回、粗精排、重排、混排、结果呈现。如图3所示。

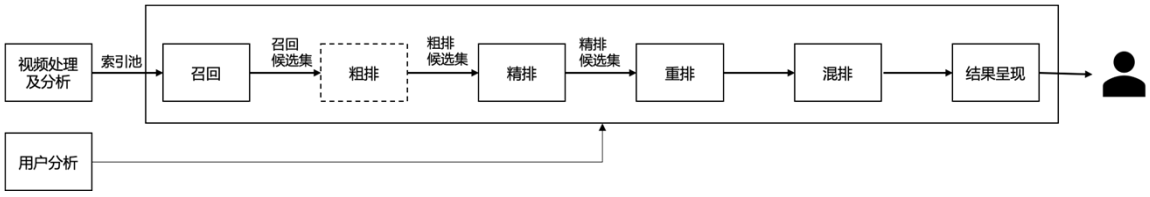

图3 个性化视频推荐流程图

其中：

a) 召回：在展示推荐场景下的索引池中，利用人工智能技术检索出满足用户偏好的视频集合，生成召回候选集。由视频召回模块完成；
b) 粗排、精排、重排：在召回候选集中，基于视频内容的理解和用户分析，完成推荐视频的确定和排序。由视频推荐决策模块中的视频排序、决策生成共同完成：
   1) 粗排使用的视频排序策略目标相对较少。粗排的性能效率相对要求较高；
   2) 精排使用的视频排序策略目标相对较多。精排的结果准确性要求较高；
c) 混排：将经过粗排、精排、重排的原生视频与其他领域视频进行混合编排。由视频推荐决策模块中的决策生成完成；
d) 结果呈现：将最终的视频推荐结果主动呈现给用户。由视频内容呈现模块完成。

参 考 文 献

[1] 中华人民共和国个人信息保护法，
http://www.gov.cn/xinwen/2021-08/20/content_5632486.htm
[2] 中华人民共和国未成年人保护法，
http://www.gov.cn/xinwen/2020-10/18/content_5552113.htm
[3] 儿童个人信息网络保护规定，
http://www.cac.gov.cn/2019-08/23/c_1124913903.htm
[4] 网络信息内容生态治理规定，http://www.cac.gov.cn/2019-12/20/c_1578375159509309.htm