ICS 35.030
CCS L 80

# 中华人民共和国国家标准

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

**Data security technology—Security requirements for automated decision making based on personal information**

2025-03-28 发布　　2025-10-01 实施

国家市场监督管理总局
国家标准化管理委员会　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。

本文件由全国网络安全标准化技术委员会(SAC/TC 260)提出并归口。

本文件起草单位：北京理工大学、中国信息通信研究院、中国电子技术标准化研究院、中国网络安全审查技术与认证中心、北京抖音信息服务有限公司、北京百度网讯科技有限公司、北京尚隐科技有限公司、北京三快在线科技有限公司、贝壳找房(北京)科技有限公司、上海美士达商务咨询有限公司、上海杉涌律师事务所、北京市竞天公诚律师事务所、北京小桔科技有限公司、北京汉华飞天信安科技有限公司、携程旅游信息技术(上海)有限公司、蚂蚁科技集团股份有限公司、阿里巴巴(北京)软件服务有限公司、北京快手科技有限公司、北京京东尚科信息技术有限公司、北京微梦创科网络技术有限公司、北京腾云天下科技有限公司、北京市中伦律师事务所、北京外国语大学、中国政法大学、北京电子科技学院、云从科技集团股份有限公司、荣耀终端有限公司、北京深度求索人工智能基础技术研究有限公司、OPPO广东移动通信有限公司。

本文件主要起草人：洪延青、田申、葛鑫、吴梦漪、朱曼莉、王劲松、张朝、赵冉冉、刘笑岑、薛晶、徐全全、万方、张凌寒、王玎、彭根、樊华、王磊、陈湉、何延哲、刘影、葛梦莹、王敬周、落红卫、孙铁、许锐、张娜、李昳婧、刘榕、顾伟、郭建领、周杨、吴佳蔚、呼娜英、丁晓强、胡立平、付艳艳、白晓媛、石玉珍、赵晓娜、李军、彭骏涛、吴少卿、黄蓉、范晔、梁天翔。

# 数据安全技术　基于个人信息的自动化决策安全要求

## 1　范围

本文件提出了基于个人信息的自动化决策的基本安全原则，规定了通用安全要求、算法安全要求、特征生成安全要求、决策安全要求和自动化决策典型场景特别安全要求。

本文件适用于开展自动化决策的个人信息处理者规范其算法开发、特征生成和决策活动，也适用于监管部门、第三方评估机构对自动化决策进行监督、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　信息安全技术　个人信息安全规范

GB/T 41391　信息安全技术　移动互联网应用程序（App）收集个人信息基本要求

GB/T 41479　信息安全技术　网络数据处理安全要求

GB/T 42888—2023　信息安全技术　机器学习算法安全评估规范

## 3　术语和定义

GB/T 35273 界定的以及下列术语和定义适用于本文件。

3.1

**自动化决策　automated decision making**

通过计算机程序自动分析和评估个人的行为习惯、兴趣爱好或者经济、健康和信用状况等，并进行决策的活动。

注：自动化决策可进一步分解为特征生成和决策两个过程。

3.2

**个人特征信息　personal characteristics information**

通过计算机程序自动处理个人信息后，并由计算机程序自动生成的关于特定自然人的偏好、职业、经济、健康、教育和信用情况等的信息。

注：个人特征信息不包括个人生物识别信息。

3.3

**个人特征信息生成　generation of personal characteristics information**

通过计算机程序自动处理个人信息，经过个人特征提取、特征选择、特征计算和特征输出等步骤，生成开展针对个人决策所需的输入信息的过程。

注：简称“特征生成”。

3.4

**决策　decision making**

以计算机程序生成的个人特征信息为输入，计算机程序产生的能影响个人自身状态、其所处的物理

或虚拟环境的预测、内容、建议和决定等输出。

注：在实践中，计算机程序的输出除了提供给个人的一般性信息内容之外，还包括雇佣或解聘特定个人、授予特定个人贷款、针对个人开展特定的信息推送和商业营销、针对服务提供报价等能对个人的人身、财产、精神等状态等产生影响的决定。

3.5

**移动互联网应用程序　mobile internet application；App**

运行在智能移动终端上的应用程序。

注：包括智能移动终端预置、下载安装的应用程序和小程序。

3.6

**对个人权益有重大影响的决定　decision with significant impact on individual's rights and interests**

对个人法定权益的实现造成法律影响以及对个人其他权益造成类似显著影响的决定。

注 1：影响包括正面或负面影响。

注 2：法律影响包括但不限于：被赋予享有或被剥夺享有一种法律上的特定权益，如子女抚养或房产权益；被剥夺或限制行动自由；拒绝进入边境或特定区域；被有关执法机构强制增加安全或监管措施等。

注 3：类似显著影响包括但不限于：对个人不合理的排斥或歧视行为、对个人可能产生长期或永久影响的决策行为，以及其他对个人的处境、行为或选择产生普遍认知方面的重大影响。

3.7

**不合理的差别待遇　unreasonable differential treatment**

通过收集和分析作为交易相对方的用户的交易信息、浏览内容及次数和交易时使用的终端设备品牌及价值等情况，对交易条件相同的交易相对方不合理地提供不同交易条件的行为。

3.8

**算法开发方　algorithm developer**

根据个人信息处理者提出的需求，开发出满足所提需求的自动化决策算法的主体。

注：业界实践中普遍存在两种情形：一是个人信息处理者委托他人开发自动化决策算法；二是个人信息处理者自行开发其所需的自动化决策算法。

3.9

**机器学习算法　machine learning algorithm**

功能单元通过学习新知识技能或整理已有知识技能以改进其性能的算法。

[来源：GB/T 42888—2023，3.1]

## 4　概述

### 4.1　自动化决策过程

自动化决策分为两个过程：特征生成和决策，见图 1。

特征生成的过程包括特征提取、特征选择、特征计算和特征输出等步骤。特征提取是指从原始的个人信息中提取出有价值信息的过程。特征选择是指从提取的特征中选择就特定业务目的来说最有用的特征的过程，其目的是降低维度，减少计算量，并避免过拟合。特征计算是指基于选择的特征计算新的特征或修改现有特征的过程。特征输出是指将处理好的个人特征信息作为输入，以支持针对个人的决策。得益于目前大数据和人工智能技术的发展，上述步骤主要由计算机程序自动开展，无需人工参与。

决策是在特征生成所提供的个人特征信息的参与下，对个人采取具体行动。决策活动可不同程度人工参与，也可无需人工参与。

注：决策基本上完全由计算机程序作出具体决定，人工的介入往往发生在特定个人对决策提出异议之后。

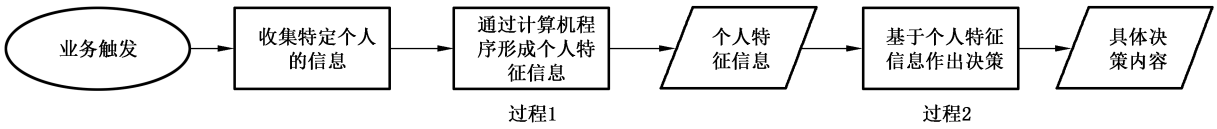


图 1　自动化决策过程

### 4.2　自动化决策使用的计算机程序及其算法

自动化决策强调通过计算机程序自动完成“对个人的行为习惯、兴趣爱好或者经济、健康、信用状况等”的“分析、评估”，突出个人信息处理者为实现特定目的而事先自行开发并部署（或使用第三方开发的）特定的计算机程序。如为了完成精准广告投放，个人信息处理者有意识地通过在网站或 App 页面上“埋点”收集个人的点击或浏览等行为信息，并通过预设的特征模型计算出特定个人特征信息。

除了特征生成由计算机程序完成之外，决策也可由计算机程序辅助人工完成，或完全由计算机程序根据特征生成形成的个人特征信息而作出。

计算机程序是具体的指令序列，而算法是对计算机上执行的指令序列背后的计算过程的具体描述。因此完成特征生成的或用于决策的算法，在很大程度上决定了自动化决策过程的透明性和自动化决策结果的公平公正性等。

### 4.3　自动化决策处理信息的范围

自动化决策所处理的个人信息包括但不限于：

a）　个人主动提供的个人信息；

b）　个人信息处理者自动收集的个人信息；

c）　个人信息处理者从第三方获得的个人信息；

d）　以前述三类个人信息为处理对象，经由计算机程序分析形成的衍生信息（如与个人相关的标签、参数等）。

### 4.4　自动化决策的安全风险

自动化决策的安全风险包括但不限于：

a）　为开展特征生成，个人信息处理者违法违规或过度收集用户个人信息、未充分告知处理目的；

b）　在特征生成过程中，个人信息处理者在不具备合法性基础前提下（如超出用户明确授权范围）对个人信息过度分析或挖掘，侵犯用户隐私或造成用户焦虑恐慌；

c）　对计算机程序开展特征生成背后的算法逻辑或模型等，无法做出清晰准确的解释；

d）　特征生成所得出的个人特征与客观情况不符或不准确；

e）　特征生成所得出的个人特征包含淫秽、色情、赌博、迷信、恐怖和暴力等内容，或表达对民族、种族、宗教、残障和疾病等歧视的内容；

f）　未经充分评估，仅根据特征生成所提供的信息而对个人开展决策；

g）　未根据自动化决策造成个人合法权益影响的程度，适当地设置人工介入、干预和修正决策结果的方法；

h）　未向个人披露开展自动化决策的具体方式或范围，损害个人的知情权、选择权和拒绝权等；

i）　在开展自动化决策之前，未依法开展事前评估而造成不公平或不公正的决策结果，或决策结果对个人权益造成不利影响且无法修正。

## 5 安全原则

开展自动化决策时，个人信息处理者按以下安全原则实施安全控制措施。

a） 合法正当原则：应符合相关法律法规规定，不应以欺诈、诱骗和误导的方式进行处理。

注：开展自动化决策的合法性基础包括充分告知，取得同意或为订立/履行个人作为一方当事人的合同所必需或履行法定职责或法定义务所必需等。

b） 公开透明原则：应以明确、易懂和合理的方式向用户公开自动化决策的方式、范围、逻辑、目的和规则等。

c） 公平公正原则：应按维护社会公平和道德伦理的精神进行处理活动，自动化决策的结果应公平和公正，不应造成不合理的差别待遇。

d） 主体参与原则：应向用户提供关于自动化决策的选择、解释、拒绝和投诉等方法。

e） 数据质量原则：自动化决策应保证所处理的个人信息的质量，避免因个人信息不准确和不完整对个人权益造成不利影响。

## 6 通用安全要求

个人信息处理者开展自动化决策符合以下要求。

a） 自动化决策相关的数据处理活动应符合 GB/T 41479 的要求。

b） 自动化决策相关的个人信息处理活动应符合 GB/T 35273 的要求，如系通过 App 开展自动化决策处理行为，还应符合 GB/T 41391 的要求。

c） 应针对自动化决策建立专门的风险管理流程，包括但不限于：

1） 加强个人信息处理全生存周期保护，在开展自动化决策处理动作前开展个人信息安全影响评估，并对处理情况进行记录，使日志记录能支持事件溯源和处置；

2） 定期自行或委托外部机构对自动化决策处理行为进行合规审计；

3） 采取加密等方式实现自动化决策开展所涉及的个人信息从生产环境到开发环境，再到线上运行的安全传输，并在处理活动中对所涉及的敏感信息进行脱敏处理；

4） 建立完整的对画像标签等数据访问权限的申请、审批、授权和回收的权限管控；

5） 为用户提供便捷明晰的意见反馈渠道，并设置专门责任人跟进处置响应。

## 7 算法安全要求

### 7.1 算法影响评估

个人信息处理者或算法开发方在自动化决策算法开发前应符合以下要求。

a） 明确自动化决策算法所用于的目的或场景，包括但不限于：

1） 使用算法的产品或服务所属的行业或领域，以及具体用户需求；

2） 使用算法的产品或服务与对象人群的互动方式；

3） 算法所适用的对象人群的基本情况和一般期待。

b） 明确自动化决策算法就所用于的目的或场景来说所支持的具体任务，如计算价格、信息内容推荐等。

c） 开展自动化决策算法影响评估。评估内容包括但不限于：

1） 预期的自动化决策算法的行为表现及相应的效率或收益的提升；

2） 明确非预期或错误的自动化决策算法的行为表现及相应的成本；

3） 根据预期的自动化决策算法的使用情况和过去在类似情况下对自动化决策算法的使用、事件报告或其他数据，分析自动化决策算法可能产生负面影响的风险，特别是对个人权益的影响。

d） 在自动化决策算法可能对个人权益造成重大影响时，可聘请第三方或组建内部专家委员会开展自动化决策算法影响评估。

e） 根据自动化决策算法影响评估的结果，明确与算法支持的具体任务相适配的算法安全目标，包括但不限于算法安全的技术特征要求、逻辑特征要求、人工介入、训练和测试数据要求。

### 7.2 算法安全技术要求

个人信息处理者或算法开发方使用的自动化决策的算法应符合以下要求：

a） 能正确捕捉训练数据中存在的相关或因果关系；

b） 能在可接受的统计误差范围内持续产生相同的结果；

c） 对不可控因素的变化具有最小的敏感性；

d） 能抵御对抗性攻击。

### 7.3 算法安全可信要求

个人信息处理者或算法开发方使用的自动化决策的算法应符合以下要求：

a） 个人信息处理者能提供一个关于算法分析或评估结果如何产生的程序性或原理性描述；

b） 算法分析或评估结果就其所用于的目的或场景来说具有能被个人所认知的含义；

c） 算法对个人私密空间、私密活动和私密信息的侵扰可度量和可控制；

d） 算法基于具有人群代表性的数据开展训练，以及算法设计中避免人为歧视性结果；

e） 算法一致地适用于就其所用于的目的或场景来说条件相同的个人。

### 7.4 算法安全人工介入要求

个人信息处理者或算法开发方应符合以下要求。

a） 具有人机交互的界面、调试工具等人工介入功能，以便个人信息处理者或算法开发方指定的管理人员能有效监督自动化决策算法及其计算机程序在运行过程中持续符合算法安全的技术特征要求和逻辑特征。

b） 使人机交互的界面、调试工具等人工介入功能能让个人信息处理者或算法开发方指定的管理人员：

1） 了解自动化决策算法的能力和限制，并监测其运行状况，以便尽快检测并解决运作异常、功能失调和突发问题；

2） 正确、完整地理解自动化决策算法的输出；

3） 在任何特定情况下决定不使用自动化决策算法及其计算机程序，或忽略、推翻或撤销自动化决策算法的输出；

4） 能干预自动化决策算法及其计算机程序的运行，尤其是中断其运行。

### 7.5 保障算法安全的训练和测试数据要求

个人信息处理者或算法开发方使用的训练和测试数据应符合以下要求：

a） 真实性、准确性、客观性和多样性符合要求；

b） 相关数据在最初选择和标注等环节不存在缺乏公平、包含偏见和歧视或范围是否与算法目标相匹配的问题；

c） 不存在被攻击或污染的情况；

d) 在用于训练和测试时具备合适的合法性基础，如样本数据采集和标注阶段、测试数据在模型测试阶段和算法模型收敛且可用于生成个人信息阶段（如使用训练好的算法模型对个人进行信用评分或偏好预测）等。

注：对于预处理（向量化）的数据，也可能会构成个人信息，若确实有证据可证明无法直接或间接根据在模型训练过程中所涉及的数据用于关联或识别个人，即相关数据进行了匿名化处理，此类数据不再构成个人信息。

### 7.6 算法开发技术文档要求

个人信息处理者或算法开发方开发自动化决策算法应形成配套的技术文档以供相关方查阅。技术文档的内容包括但不限于以下内容。

a) 对自动化决策算法的一般描述，包括：
   1) 预期支持的具体任务、系统开发人员、日期和系统版本；
   2) 自动化决策算法与系统外的硬件或软件交互方式；
   3) 相关计算机程序或固件的版本以及所有版本更新要求；
   4) 对自动化决策算法投入使用的所有形式的说明；
   5) 对使用自动化决策算法的硬件的描述；
   6) 使用自动化决策算法的产品的外部特征照片或插图、标记和内部布局；
   7) 使用说明或安装说明。
b) 对自动化决策算法的要素和开发流程的说明，包括：
   1) 为开发自动化决策算法所采取的方法和步骤，包括使用第三方提供的预训练系统或工具以及提供方使用和集成或修改这些系统或工具的方式；
   2) 自动化决策算法的设计规范，即通用逻辑、关键设计选择（包括所作的理由和假设，也涉及该系统拟针对的个人或群体）、主要分类选择和算法旨在优化的内容，以及不同参数的相关性；
   3) 为实现算法的技术特征要求和逻辑特征要求而采用的技术解决方案相关的任何可能的权衡决定；
   4) 对计算机程序架构的描述，说明程序组件如何相互依赖或相互输入并集成到整体处理中；
   5) 用于开发、训练、测试和验证自动化决策算法的计算资源。
c) 对自动化决策算法训练过程的记录，特别是对训练方法和技术以及使用的训练数据集进行描述（包括：数据集的来源、范围和主要特征、获取和选择数据的方式、标签程序和数据清理方法等）。
d) 使用的验证和测试程序的相关信息，包括：
   1) 所用验证和测试数据及其主要特征的信息；
   2) 用于衡量对算法技术特征要求和逻辑特征要求的符合性的指标；
   3) 测试日志以及所有由责任人注明日期并签名的测试报告。
e) 对人工介入要求所采取措施的实施评估情况。

### 7.7 算法安全运行要求

个人信息处理者应符合以下要求：

a) 明确针对技术特征、逻辑特征和人工介入要求的定量或定性监测指标和监测方法；
b) 明确对应各个负面影响的定量或定性监测指标和监测方法，如用户反馈、投诉和事件舆情等；
c) 建立有效的影响追踪方法，在必要的情况下，通过外部人员协助监测自动化决策算法及其计算机程序与预期行为表现的偏差，如功能蠕变；
d) 基于监测结果定期开展算法影响评估，对个人权益有重大影响的自动化决策算法应每半年开展一次算法影响评估；

e) 根据算法影响评估结果采取减轻负面影响的措施；
f) 事先建立负面影响应对方案和措施，以更新、替代或停用与其预期行为表现不一致的自动化决策算法及其计算机程序。

### 7.8 其他要求

如个人信息处理者或算法开发方使用的是机器学习算法，应符合 GB/T 42888—2023 中 5.1.2～5.1.4 的要求。

## 8 特征生成安全要求

### 8.1 特征生成的个人信息处理要求

#### 8.1.1 基本要求

个人信息处理者应选择和收集与特定业务目的相匹配的个人信息类型。

#### 8.1.2 个人信息真实性要求

个人信息处理者应采取有效措施提升所收集的个人信息的真实性，符合的要求包括但不限于：
a) 确认个人信息来源真实，即保障个人信息来自原始、可靠的数据源，而非虚构或伪造的来源；
b) 确认个人信息未经篡改，即保障个人信息在收集、处理和传输的过程中保持了原始状态，没有受到恶意篡改、损坏或删除；
c) 必要的情况下开展身份验证，即保障个人信息提供者或收集者的身份是否经过验证，其身份情况是否能保障个人信息的真实性和准确性；
d) 确认个人信息收集的透明度，即保障个人信息收集和清洗的过程透明，以及可审查和可验证，以保证个人信息的真实性。

#### 8.1.3 个人信息准确性要求

个人信息处理者应采取有效措施提升所收集的个人信息的准确性，符合的要求包括但不限于：
a) 确认所收集的个人信息类型应能反映个人信息主体的实际情况，必要时可采取相关措施对个人信息的客观性进行验证；
b) 确认所收集的个人信息的相关性和典型性；
c) 确认所收集的个人信息为最新状态，避免因信息过时而影响个人信息的准确性。

注：如果数据不准确或过时，依赖于此的特征生成也可能由此建立在过期的数据或带有错误解释的外部数据基础之上。

#### 8.1.4 收集个人信息合法性要求

个人信息处理者收集个人信息应具备合法性，符合的要求包括但不限于以下内容。
a) 根据收集个人信息的不同阶段以及不同场景考虑选择适当的合法性基础，包括但不限于：
   1) 基于同意时，提供充分清楚且全面的信息使个人信息主体理解其同意的内容，引入颗粒化同意流程，提供简单的路径为不同目的的数据处理获取同意并提供便捷的撤回同意方法；
   2) 基于履行合同之必要时，根据具体场景下的合同或服务目的的必要性和正当性适用，且客观上已衡量过其他的方式后谨慎采用；
   3) 基于履行法定义务时，仅在出于公共利益、风险防控或保障个人信息处理者提供服务的安全性和可预期性等目的，并制定适当的措施保障用户的权利和正当利益后开展，如预防欺

诈、反作弊和反洗钱等情形。

b） 在最小范围和数量内收集为实现特征生成所必需的个人信息。

### 8.2 特征生成的计算安全要求

#### 8.2.1 基本要求

个人信息处理者使用的算法应符合第7章的要求。

#### 8.2.2 特征提取要求

个人信息处理者应：

a） 预先对所收集的个人信息进行有效的预处理，包括数据清洗和特征工程等；

b） 排查所收集的个人信息的缺失值、异常值和重复值，规范个人信息的格式、结构和类型等要素；

c） 提取的数据具备一致性、有效性和可用性；

d） 提取的数据特征不存在偏见和歧视等情形。

#### 8.2.3 特征选择和特征计算要求

个人信息处理者应：

a） 选择对实现特定业务目的最有预测能力的特征；

b） 考虑到选择的特征之间的相关性，避免过多的相关性冗余；

c） 避免选择可能引入不公平或歧视性待遇的特征；

d） 保证选择数据特征时的平衡性，避免因偏重或缺失某一类型的个人信息导致特征计算和输出存在错误的情形；

e） 开展进一步的特征计算，如发现选取的数据特征与实现特定业务目标之间存在不足关系时。

#### 8.2.4 特征输出要求

个人信息处理者应：

a） 保证输出的对个人信息主体特征的描述不表达对性别、民族、种族、年龄、宗教、残障、疾病和性取向歧视的内容；

b） 对特征生成所使用计算机程序中的函数、数据和模型等进行审慎评估和测试，并采用特定的方法削弱程序输出可能带来的偏见和歧视，保障输出的可解释、可理解、准确性、非歧视性和公平性。

## 9 决策安全要求

### 9.1 基本要求

个人信息处理者基于特征生成进行针对个人的决策时：

a） 不应侵害公民、法人和其他组织的合法权益；

b） 不应危害国家安全、荣誉和利益，煽动颠覆国家政权和推翻社会主义制度，煽动分裂国家和破坏国家统一，宣扬恐怖主义和极端主义，宣扬民族仇恨和民族歧视；

c） 不应传播暴力和淫秽色情信息，编造和传播虚假信息扰乱经济秩序和社会秩序。

### 9.2 决策前的告知要求

个人信息处理者在使用自动化决策程序之前，应向个人提供关于自动化决策的相关信息，包括但不

限于以下内容。

a） 易于理解的自动化决策处理说明，包括但不限于：

1） 自动化决策处理动作所基于的个人信息；

2） 自动化决策处理所涉及的特征生成和决策的原理。

b） 个人对自动化决策行为及结果享有的权利和行使权利的方式。

c） 便捷有效的反馈渠道并处理反馈意见。

注：个人信息处理者可在向个人提供的个人信息处理规则中说明上述内容，也可向个人提供单独的说明文档。

### 9.3 决策中的个人权益保障要求

个人信息处理者应基于特定的自动化决策场景为个人提供权益保障，包括但不限于以下内容。

a） 响应解释请求。在响应个人对自动化决策的解释请求时，针对提出解释请求的个人的基本情况，在不影响或损害个人信息处理者商业秘密或其他合法权益的情况下，适当地说明决策逻辑、价值权重和个人信息利用情况等。

b） 响应干预请求，包括但不限于：

1） 向个人提供选择或删除用于算法推荐服务的针对其个人特征的标签的功能或途径；

2） 在通过自动化决策方式向个人进行信息推送和商业营销时，向个人提供不针对其个人特征的选项，或向个人提供便捷的拒绝方式；

3） 在作出对个人权益有重大影响的决定时，向个人提供拒绝个人信息处理者仅通过自动化决策的方式作出决定的方法或途径。

## 10 自动化决策典型场景特别安全要求

### 10.1 教育或职业机会

在教育、工作机会推荐或候选人评估过程中通过自动化决策分析人选特征，并根据特定特征指标完成筛选匹配时，个人信息处理者：

a） 不应设置歧视性或偏见性用户标签并据此设定筛选策略，如仅在通过行为分析推测候选人为中年已育女性的情况下，直接将该候选人筛出面试名单，或设置为其匹配高薪主管岗位的机会远低于同等情况下的男性候选人的系统策略等；

b） 应为用户提供拒绝个人信息处理者仅通过自动化决策的方式作出决定的方法或途径。

### 10.2 信用贷款或保险评估

在判定是否向某位申请人发放贷款或通过某项保险险种申请、确定贷款发放额度等场景下，通过自动化决策分析人选行为特征推断其信用、健康情况作出决定时，个人信息处理者：

a） 不应在用户不知情且未获得用户授权或具备其他合法性基础前提下，仅通过对该用户的 App 浏览行为习惯等间接指标，或者使用与用户具有同类特征的用户群体特点间接推断得出信用评分，作出不通过其前述申请的决定；

b） 不应设置歧视性或偏见性用户标签，如残障和重大疾病等标签内容，如评估行为所必需，仅得如实、客观和准确反映该用户身体特征及疾病情况；

c） 应为用户提供拒绝个人信息处理者仅通过自动化决策的方式作出决定的方法或途径。

### 10.3 社会福利资格等公共治理领域

通过自动化决策在行政管理活动中作出如给予或不给予行政许可或审批等决定的行为时，个人信息处理者应：

a) 考虑到决策行为可能对个人产生重大影响，提供便捷有效的方式为个人提供申诉渠道，及时对个人诉求予以处理；
b) 建立安全可行的公示方式，在合理范围内保障公众知情权，接受公众监督；
c) 为个人提供拒绝个人信息处理者仅通过自动化决策的方式作出决定的方法。

### 10.4 劳动关系领域

针对劳动者进行自动化决策处理时，个人信息处理者：

a) 应公示相关自动化决策算法的机制、机理，如对配送服务从业者公示配送时间预估、路线规划、配送费用计算明细等相关算法机制机理；
b) 不应利用自动化决策算法对劳动者进行压榨和操纵；
c) 不应侵害法律法规所规定的劳动者合法权益；
d) 利用算法进行绩效管理、人事管理时，应在算法设计层面关注劳动权益保护；
e) 向劳动者提供工作调度功能时（包括劳动报酬、休息休假、订单分配、工作时间和奖惩措施等功能），应为劳动者设立便捷明晰的意见反馈和投诉处理方法或途径。

### 10.5 特殊群体的自动化决策安全要求

#### 10.5.1 儿童

针对儿童进行自动化决策处理时，个人信息处理者：

a) 除有充分事由且考虑了儿童的最大利益并采取适当措施保护儿童免受侵害外，原则上不应对儿童开展自动化决策处理；
b) 如需对儿童进行自动化决策（如为了保护儿童的福利），需要提供适合该年龄段理解的信息，说明儿童个人信息会怎么处理以及潜在风险，并采取适当的保障措施有效保护儿童的权利和合法利益；
c) 不应利用算法推荐服务诱导儿童沉迷网络；
d) 不应通过自动化决策方式向未成年人进行商业营销。

#### 10.5.2 老年人

在对移动终端、App 等开展适老化改造时，个人信息处理者应：

a) 根据老年人使用习惯和权益保障需要开展自动化决策；

注：适老化改造方法见 GB/T 37668—2019 的相关内容。

b) 持续优化完善面向老年人的算法推荐服务，便利老年人获取有益身心健康的信息。

### 10.6 信息推送、商业营销

利用特定个人或其所在群体相关的个人信息通过自动化决策方式向个人进行个性化内容推送时，如信息推送或商业营销等，个人信息处理者：

a) 不应强制要求用户选择兴趣标签，允许用户跳过标签选择页面；
b) 应提供兴趣标签查看功能，向用户展示用于个性化内容推送的个人兴趣标签；
c) 应向用户提供便捷的关闭个性化内容推送的选项，用户选择关闭后，平台应立即停止个性化内容推送且不影响用户正常使用，不应频繁通过弹窗等方式提醒用户开启；
d) 不应利用用户标签进行诱导营销和过度推荐；
e) 不应将包含炫富、色情、暴力、极端和低俗等违法和不良信息关键词记入用户标签，并向其推送个性化内容；

f) 不应设置歧视性或偏见性用户标签，并向其推送个性化内容；
g) 不应实施算法屏蔽信息、过度推荐、操纵榜单和控制热搜等可能造成用户信息茧房的操纵行为；
h) 应通过内容去重、打散干预等策略提升个性化内容推送的多样性、丰富性；
i) 应健全针对水军刷榜、水军账号等违规行为的账号检测识别技术手段，严格管控不法分子恶意利用榜单排序规则操纵榜单、炒作热点行为；
j) 应通过设置“不感兴趣”“此类内容过多”“重复推荐”等功能选项，或向用户提供个人兴趣标签管理功能等形式，允许个人动态调整个性化内容推送中不同种类内容的比例。

### 10.7 商业交易

利用特定个人或其所在群体相关的个人信息向个人进行商品、产品服务营销推荐时，个人信息处理者：

a) 不应利用用户年龄、职业、消费水平等个人或群体特征，开展任何可能造成交易价格、交易机会、交易条件等实际差别待遇的行为；
b) 应提升优惠促销透明度，清晰说明优惠券的领取条件、发放数量和使用规则等内容；
c) 不应实施虚构原价、虚假优惠折价等不正当价格行为；
d) 不应对消费者收取未予以标明的费用；
e) 不应通过利用算法操纵中奖概率、中奖结果和中奖人员等欺骗方式进行有奖销售；
f) 不应滥用市场支配地位，利用算法在无正当理由的情况下操纵价格，排除和限制市场竞争。

# 参 考 文 献

[1] GB/T 37668—2019 信息技术 互联网内容无障碍可访问性技术要求与测试方法

[2] GB/T 39335—2020 信息安全技术 个人信息安全影响评估指南

[3] ISO/IEC 29100:2024 Information technology—Security techniques—Privacy framework

[4] EU General Data Protection Regulation,2016

[5] 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)

[6] 中华人民共和国电子商务法(2018 年 8 月 31 日第十三届全国人民代表大会常务委员会第五次会议通过)

[7] 中华人民共和国网络安全法(2016 年 11 月 7 日第十二届全国人民代表大会常务委员会第二十四次会议通过)

[8] 网络数据安全管理条例(2024 年 8 月 30 日国务院第 40 次常务会议通过)

[9] 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、工业和信息化部、公安部、国家市场监督管理总局发布)

[10] 生成式人工智能服务管理暂行办法(2023 年 7 月 10 日国家互联网信息办公室、中华人民共和国国家发展和改革委员会、中华人民共和国教育部、中华人民共和国科学技术部、中华人民共和国工业和信息化部、中华人民共和国公安部、国家广播电视总局令第 15 号)