

黑 龙 江 省 地 方 标 准

DB 23/T XXXX. 7—XXXX

政府网站建设管理规范 第 7 部分：集约化
平台内容安全审核

征求意见稿

联系人：黑龙江省数字经济研究会 李健航
联系电话：15124541881
邮箱：hljsszjjyh@people-ai.cn

在提交反馈意见时，请将您知道的相关专利连同支持性文件一并附上。

XXXX – XX – XX 发布

XXXX – XX – XX 实施

黑龙江省市场监督管理局 发 布

目 次

1 范围 1

2 规范性引用文件 1

3 术语和定义 1

4 总体框架 2

5 审核标准 6

6 技术对接要求 14

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。

本文件是《政府网站建设管理规范》的第7部分。《政府网站建设管理规范》已经发布了以下部分：

请注意本文件的某些部分可能涉及专利。本文件的发布机构不承担识别专利的责任。

本文件由××××提出。

本文件由××××归口。

本文件起草单位：

本文件主要起草人：

政府网站建设管理规范 第7部分：集约化平台内容安全审核

1 范围

本文件规定了黑龙江省人民政府网站集约化平台内容安全审核的要求，包括总体框架、内容安全审核要求、技术对接要求等。

本文件适用于省级、地市级人民政府网站集约化平台内容的安全审核。

2 规范性引用文件

本文件没有规范性引用文件。

3 术语和定义

下列术语和定义适用于本文件。

3.1

政务新媒体

各级行政机关、承担行政职能的事业单位及其内设机构在微博、微信等第三方平台上开设的政务账号或应用，以及自行开发建设的移动客户端等。

3.2

不良违法信息

计算机信息系统及其存储介质中存在、出现的，以图像、文字，声音、视频等多种形式表示的，含有攻击人民民主专政、社会主义制度，攻击党和国家领导人，破坏民族团结等危害国家安全内容的信息；含有宣传封建迷信、淫秽色情、凶杀、教唆犯罪等危害社会秩序的内容等。

3.3

信息系统

政府网站集约化平台建设所涉及的互动交流系统、全媒体内容管理系统、统一信息资源库系统、集约化服务系统、政务信息公开系统等应用系统。

3.4

政府数据

政府网站在信息发布、解读回应、办事服务、互动交流等网站应用在运行过程制作或获取的，以一定形式记录、保存的文件、资料、图表和数据等各类信息资源，包括政府部门直接或通过第三方依法采集的、依法授权管理的和因履行职责需要依托政府网站形成的信息资源等。

3.5

政府网站

各级人民政府及其部门、派出机构和承担行政职能的事业单位在互联网上开办的，具有信息发布、解读回应、办事服务、互动交流等功能的网站。

3.6

政府门户网站

县级以上各级人民政府开设的政府门户网站。

3.7

部门网站

省级、地市级政府部门开设的部门网站。

3.8

机审

机器审核，通过AI算法对劣质内容进行识别、过滤的一种审核模式，通过提取海量数据样本的特征，输入机器学习形成的机器算法。

注：机器审核通常可以区分为机审规则和机审模型，机审规则是最简单的机器审核方式，仅支持识别文本类内容。

机审模型则能力更加强大，可支持文本、图片、音视频的审核。

3.9

机审规则

由海量的风险词和匹配规则构成的风险词表。

3.10

标签体系

按照统一规则对内容或者用户进行分类后形成的体系。

3.11

功能模块

数据说明、可执行语句等程序元素的集合，单独命名的可通过名字来访问的过程、函数、子程序或宏调用。

注：功能模块化是将程序划分成若干个功能模块，每个功能模块完成了一个子功能，再把这些功能模块总起来组成一个整体。以满足所要求的整个系统的功能。

3.12

智能识别

利用模式识别和图像处理的技术，根据需求来自动分析和抽取原始数据中的关键信息。

4 总体框架

4.1 业务流程规范

4.1.1 审核业务规范

审核业务流程包含“人工审核”、“专家审核”、“智能机审”，详见图1：

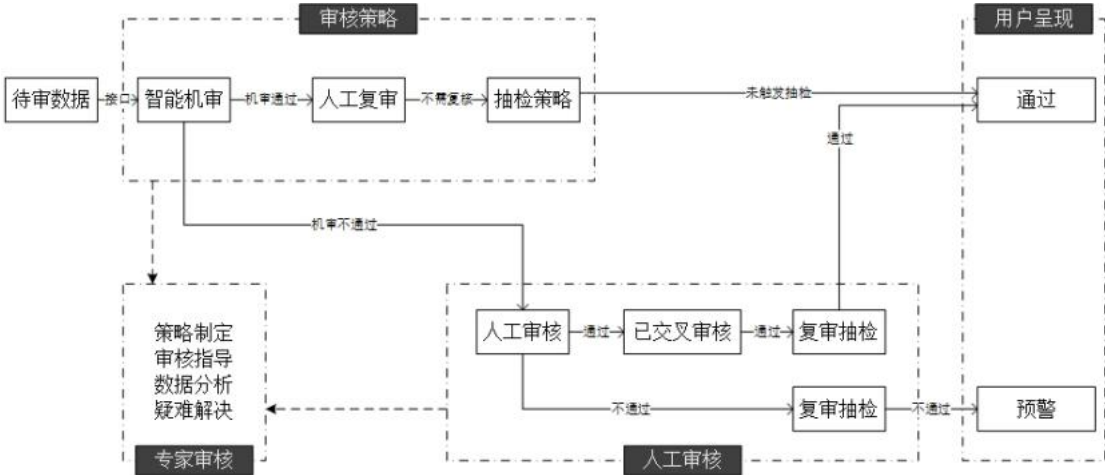


图 1 审核业务流程图

4.1.2 巡检业务规范

巡检业务是信息发布后保证内容安全工作的重要组成部分,应具备采集政务新媒体矩阵发布内容的能力,同时应具备图、文、音、视全模态数据识别能力,及时发现政务新媒体矩阵不合规信息、政府网站内被恶意篡改以及新法规发布前合规发布后不合规的内容,并如图2所示将内容及时推送预警至工作人员。

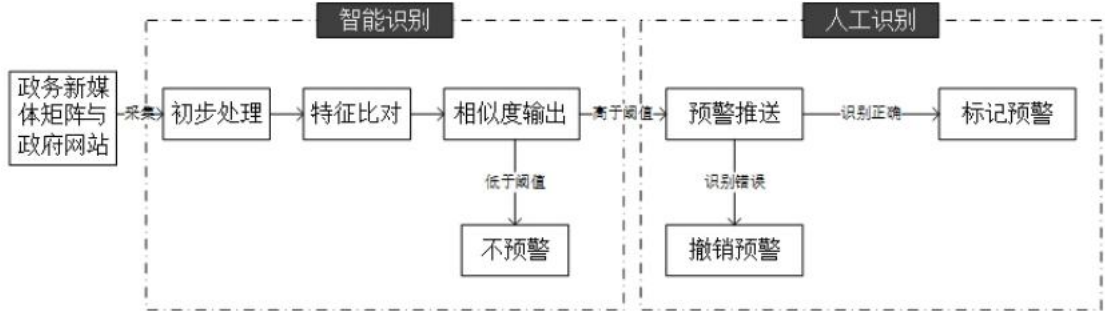


图 2 巡检业务流程图

4.1.3 业务性要求

4.1.3.1 政府网站内容安全

应形成“发布前审核”、“发布后巡检”业务闭环,政府网站的信息发布前使用智能机审与人工审核相结合的方式完成审核流程。信息发布后周期性对信息进行全量巡检,保障信息发布后及时发现被恶意篡改的内容,以及符合新颁布的法律法规等。

4.1.3.2 政务新媒体矩阵内容安全

不在发布前信息审核范围内的媒体、自媒体发布的内容应定向采集巡检,以保障内容的安全性。

4.1.4 功能性要求

4.1.4.1 人机结合审核功能模块

应具备人审、机审结合的管理功能模块，支持审核流程自定义、支持审核数据管理，支持审核员对待审数据的通过、不通过、标记等功能，支持不同用户角色管理。

4.1.4.2 文本智能识别功能模块

应具备文本智能识别功能模块，使用深度学习、上下文语义识别、NLP技术、行为识别、文本聚类、内容变种识别等技术，结合多场景、多维度判定，定位文本中有问题的部分。支持对图片/视频中的文字位置进行精准定位，并支持对楷体、宋体、仿宋、隶书、行楷、手写体、竖排文字和倾斜角度的文字检测。支持策略配置、精细化分类以及自定义词库，可高效检测涉政、暴恐、违禁、色情、谩骂等内容。

4.1.4.3 图片智能识别功能模块

应具备图片智能识别功能模块，采用深度学习、OCR、人脸识别、图片聚类、二维码检测、行为识别等技术，构建图片识别引擎，精准识别图片中落马官员、失德艺人、劣迹艺人、政治敏感内容、涉恐、涉黄、欺诈广告等红线违规违法内容。

4.1.4.4 音频智能识别功能模块

应具备智能音频识别模块，采用深度学习、普通话/英文声音转换为文字等语音识别技术。实现将数据与库中的参考模板进行比较，采用判决的方法找出最接近语音特征的能力，得出识别结果。同时支持声纹识别、娇喘、呻吟、ASMR、领导人声纹检测等功能，辅助音频检测判定结果能够支持包括WAV、MP3、WMA、FLAC、AMR、OPUS、M4A、AAC等格式，可对其进行多场景、多维度地检测与识别。如监测红线违规识别、色情、辱骂等。

4.1.4.5 视频智能识别功能模块

应具备视频智能识别模块，基于深度学习的人工智能技术，支持对视频截帧留存、识别其中包括视频画面、语音、字幕、标题中出现的红线违规等内容，例如政治内容、落马官员、色情污秽、暴恐宗教、低俗辱骂、广告水印、吸烟赌博场景等类型，支持格式包括支持MP4、AVI、MKV、MOV等。

4.1.4.6 特征库管理模块

通过建立敏感特征库，可以统一对敏感人物、敏感图像、敏感音频信息进行管理编辑，支持手动添加与删除，支持自定义人像库上传及审核，支持敏感图像标注上传。对已审核完成的人像、音频及图像标注数据，可支持实时更新到关联到相关任务的特征库。

4.2 省级与地市级平台联动

4.2.1 省市联动架构

省市联动架构图见图3：

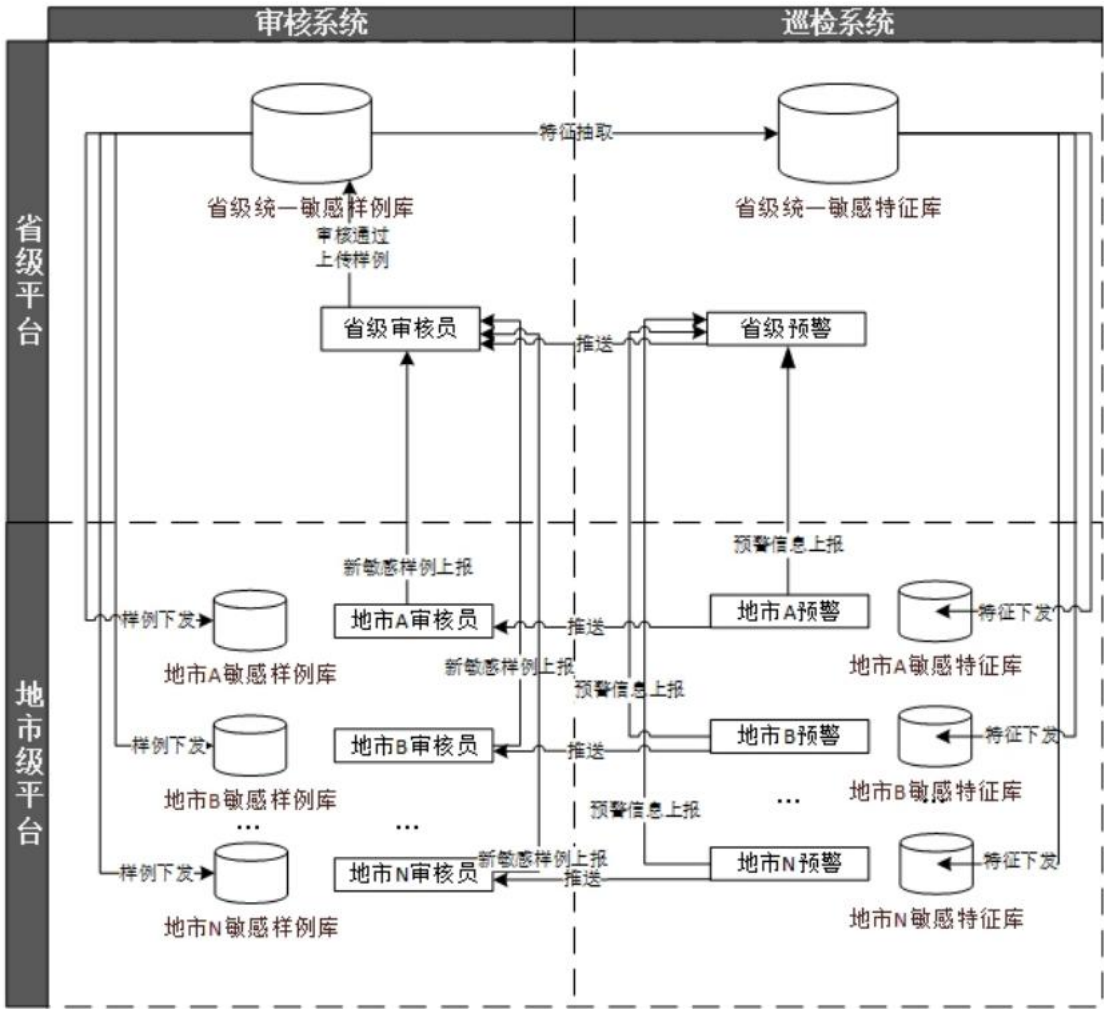


图 3 省市联动架构图

4.2.2 内容审核系统省市联动规范

4.2.2.1 敏感样例库省市联动

应在省级平台建立敏感样例总库，且作为敏感内容判断的唯一标准，同时通过接口将样例数据下发到地市单位使用，敏感样例库主要用于审核人员之间的知识传播、支撑知识图谱运行以及与巡检系统的敏感特征库联动。

4.2.2.2 审核业务省市联动

应在敏感样例库支撑的基础上建立审核业务省市联动机制，如地市在工作中遇到难以判别的实例可将该实例提交至省级平台专家协商处理，省级专家处理审核后一方面将协助结果发至地市，另一方面将该案例作为新的敏感规则上传至敏感样例库，并将该样例同步下发到其余地市。

4.2.2.3 联动权限分配

集约化平台包含省市两级平台，应建立用户角色以控制行政权限，角色包含省级专家审核员、省级审核员、市级专家审核员、市级审核员等。不同的角色可看到不同的数据量，省级用户拥有全省数据，市级用户只能查询本地市的数据，同时不同角色也在省市联动流程中负责不同的岗位职责。

4.2.3 内容巡检系统省地联动规范

4.2.3.1 敏感特征库省市联动

应在省级平台建立敏感特征总库，作为全省巡检敏感内容判断的唯一标准，同时接收来自敏感样例库的新敏感样例，将其特征化后分发至各地市凭条，同步应用至巡检任务中。

4.2.3.2 预警统计省市联动

应建立巡检监控预警和统计省市联动，各地市平台应将本地政务媒体矩阵的账号、发文数量、发文质量以及预警内容推送至省级平台，方便省级平台掌控各地市政务媒体矩阵发文质量。

4.2.3.3 联动权限分配

应建立用户角色以控制数据查看权限，省级用户可查看全省预警数据，市级用户只能查看本地市预警数据。

5 审核标准

5.1 文本审核规范要求

5.1.1 关键词过滤

能够通过与或表达、正则表达的方式配置组合关键词策略，可快速、精准定位色情、暴恐涉政、广告、辱骂等文本垃圾。

5.1.2 文本分类识别

支持业界监督学习算法朴素贝叶斯、SVM、逻辑回归、自然语言识别等技术，自动识别文本不良信息类别，包括涉黄类、政治类、商业广告类、欺诈类、涉黑类等。

5.1.3 文本聚类识别

支持采用文本编辑距离、K-means聚类等算法，自动对现网海量文本数据进行聚类，结合预置合法性类别，有效判定文本是否存在违规风险。

5.1.4 文本违规实体识别

能够将违规文本中的关键词识别定位，得到违规实体和实体位置以及实体所属的违规类型（如赌博、色情、涉政、违规人员、违规地点等）。文本违规实体间关系抽取：通过关系抽取技术，得到识别到的违规实体之间的关系，构建三元组（实体-关系-实体）。

5.2 图片及视频审核规范要求

5.2.1 涉政图片及视频识别

支持敏感专项、严格涉政、时事报道、领导人相关、英雄烈士相关、邪教迷信、落马官员相关、热点舆情、涉政综合等精细化子分类返回。对于涉政图片及视频，尤其是敏感人物/政治人物、特定场景、特定事件、特定着装等关键内容。也采用基于类似深度学习的方法，根据视频、图像信息的特定特点，进行涉政识别服务。该服务主要分为政治敏感人物识别、涉政场景识别和特定事件识别三大模块。最后，由三大模块的识别结果共同判定，得到该片段是否涉政的最终结论。

5.2.2 暴恐图片及视频识别

支持对包含暴力、血腥场景、恐怖组织头目、旗帜、涉政等涉嫌违禁的暴恐类图片及视频信息检测。能同时识别出图片中含有的暴恐元素，如：火光、刀具、鲜血、旗帜、军队等，能够根据不同场景管控需求不同，可针对性的配置，提升识别的准确度。

在采用深度学习算法，采用深度学习算法，根据截帧画面、声音、文字、视频等多维度综合决策视频结果，可快速对图片视频内容进行判别并快速定位暴恐旗帜、人物和场景以及敏感政治人物。最终在结果中返回证据画面，协助播控人员判断。下面将分暴恐和涉政两个方面进行阐述。

针对暴恐图片及视频，该服务需要对媒体内容进行深度分析，包括敏感文字识别、敏感图像识别、敏感物品检测、敏感音频识别、敏感人物检测等内容。重点关注旗帜\横幅、特定事件、行刑/血腥场面、武器、特定着装、暴乱场面以及战争场面等特定事物与场景。

5.2.3 违禁和广告图片及视频识别

支持违禁品、血腥场景、违禁人物、违禁Logo、公职服饰、不文明行为、违禁场景、火焰、骷髅、货币、毒品等细分场景识别；支持含二维码、色情广告、推广广告违规等细分场景识别；相比普通场景而言，图片视频中的不良场景兼具通用性与独特性，即不良场景是在普通场景中发生了不良状况形成。鉴于普通场景已经拥有大量数据，但不良场景数据极少，可采用迁移学习思想将普通场景数据进行很好的建模后，再迁移到不良场景中使用。迁移学习是机器学习的重要分支。近几年，迁移学习发展十分迅猛，很多学者开始挖掘深度学习模型的迁移能力。本项目采用基于模型的迁移学习——残差迁移网络来检测不良场景。

5.2.4 涉黄图片及视频识别

支持色情传播、色情性器官、色情挑逗、色情低俗段子、色情性行为、色情舆情事件、色情交友类等精细化子分类返回。该部分通过深度学习算法和实时更新的亿级图像样本库，可实现对图片与视频进行识别以及色情程度量化。视频检测采用截帧画面、声音、文字多维度综合决策视频结果，最大限度避免因截图模糊而导致误判。支持色情图片、视频的识别服务，支持批量对图像的色情程度实时打分，分数越高被判为色情图像的概率越高。其内容主要根据以下标准区分：

- a) 色情：全裸、露点、露臀（包含类似全裸）、性爱姿势、高仿真3D动画类，高仿真器具及不雅姿势（不雅姿势包括各种带有引诱性的动作）；
- b) 性感：性感动作，布料稀少衣服遮挡敏感部分大于三分之二并没有不雅姿势（如泳衣、正常哺乳类未露点等）；
- c) 正常：非色情，正常的物品、风景、人物、表情及动作。

最后，集成多种特征实现对图像的描述，采用瀑布型网络结构实现对色情图片及视频的检测。

5.2.5 图片文字转换

支持OCR，可以提取图像中包含的文字内容，再结合文本识别算法进行违规图片精准判定。OCR图文识别服务算法提供上亿字符样本积累，可精准定位图片中文字位置，准确识别斜排字，艺术字等字体。可用于字幕识别、印刷文本识别等场景。

图文识别服务应当能将电子设备中拍摄到的图片进行一系列预处理，通过深度学习的方式对图片中的文本框进行检测和识别，确定其文本的具体位置和方向，然后进行文本框的切割和处理，使其能进行文本框中单独字符的识别，深度学习的过程是对文本框的确定起着至关重要的作用，所以我们在进行深度学习的时候不断的进行模型的训练和性能的提升，在预测出所有的文本框后和图片进行比对，计算识别率和准确率，并对此进行不断的改进，以适应项目中的所有需求。

在通过深度学习完成图片中文本框的检测之后，就需要进行OCR的单独字符识别，在识别的过程中会调用识别算法的接口，将文本框中的文字根据其角度进行文本的矫正和重定位，之后在样本库中进行文字的对比识别，将文本框中的所有字符检测识别，再根据特定的组合，将图片中所有的信息输出。

5.3 音频审核规范要求

5.3.1 声纹识别检测

支持利用声纹检测进行娇喘声识别，精准识别娇喘、呻吟、耳骚等无语义的违规声音，高效过滤违规语音信息；声纹识别模块是对语音数据中的说话人身份信息的分析和处理，负责将语音数据中说话人身份信息提取出来并建立声纹模型，通过对语音中说话人身份信息和声纹库中的声纹模型进行相似度匹配，来判断当前说话人的身份；同时也可以找出多条语音或多个声纹模型中具有相同说话人身份信息的数据。

5.3.2 语音内容识别

语音转写模块是把语音数据转换为文字信息，语音识别模块借助于声学模型和语言模型，将语音信号转换为文字信息。而文本翻译模块则负责将一种语言下的文字转换为另外一种语言，如将英文转换为中文。基于自然语言处理技术、语音特征提取等多项能力，有效识别音频中涉黄、涉政、暴恐、违禁、广告等多维度的违规内容，高效过滤违规不良语音内容。

5.3.3 音频智能识别

智能音频过滤采用基于ffmpeg的音频信息动态转码技术、基于深度学习的语音识别技术和智能特色语义分析技术，能够对绝大部分音频格式进行多场景、多维度地检测与识别。其中包括落马官员、失德艺人、劣迹艺人、涉政、涉黄、广告导流等诸多类型。智能音频过滤产品识别准确率高，处理速度快，检测场景广泛，能够极大降低人工审核成本，提升音频内容质量。

5.3.4 色情污秽违规识别

运用海量的实时更新的敏感词库和文本色情模型，识别音频中涉及性器官、性行为、色情交易、娇喘等无语义的色情内容或引诱性暗示的内容，回调给客户出现的位置和命中的敏感词信息。

5.3.5 红线违规识别

运用海量的实时更新的敏感词库和文本涉政模型，识别音频中的落马官员、失德艺人、劣迹艺人、国内外领导人、宗教人物、英烈人物、封杀人物、政治机构、敏感政治事件、危险违禁品、涉赌等言论，并回调给客户出现的位置和命中的敏感词信息。

5.3.6 谩骂、不文明用语违规识别

运用海量的实时更新的敏感词库和文本辱骂模型，识别音频中含有人称和性器官的辱骂或口头禅式的不文明用语，回调给客户出现的位置和命中的敏感词信息。

5.3.7 广告内容违规识别

运用广告模型和行为策略，识别音频中含有引诱交易的涉嫌广告的描述以及有广告行为的账号，回调给客户出现的违规内容的位置。

5.3.8 性别识别

运用成熟的性别模型，识别音频中说话人的性别。

5.4 文档审核规范要求

能够提供对常见文档的审核能力，精准识别文档中涉政，涉黄，涉暴，广告违禁，低俗，公众事件等多种违规内容。

文档解析模块：接口直接传入文档，须支持多种文档格式解析，智能解析文件中的文本和图片，检测并过滤文件中的有害信息，可支撑业务各项badcase的多次检测核验。

文档支持txt、doc、docx、ppt、pptx、xls、xlsx、pdf等格式。

5.5 审核结果规范

5.5.1 规范用语

参考并依据出版行业标准规范或权威资料，结合人工智能技术发现中文错别字、多字、少字、人名、地名、英文（单词拼写）、部分成对标点、科技计量、重句、异形词、也作、曾用词等各类错误；领导人姓名、职务和领导人排序错误。

5.5.1.1 内容及知识检查

易错词检查、敏感内容检查、重点词检查、领导人信息检查、英文拼写检查、译文检查、国际单位符号检查、不规范名词检查、标点符号检查、异体字检查、繁简误用、纪年检查、地名检查、古诗文检查；

5.5.1.2 体例及格式检查

图/表/公式序号及引用关系检查、大纲标题检查、数学元素符号及体例检查、参考文献检查、脚注、尾注参见落空检查、列表序号检查、千分位、全半角、单位间隙检查；

5.5.1.3 词汇依据

词汇依据来源详细信息见表1

表 1 词汇依据来源信息

依据	版本/标准号	出版社/发布单位	作者	主管部门	归口单位
《作者编辑常用标准及规范》	第四版	中国标准出版社	中国标准出版社	—	—

《标点符号用法》	GB/T 15834-2011	中华人民共和国国家质量监督检验检疫总局、中国国家标准化管理委员会	—	教育部（语言） （360-2）	全国语言文字标准化技术委员会 （TC500）
《图书编辑校对实用手册》	第五版	广西师范大学出版社	黎洪波，利来友	—	—
《通用规范汉字表》	第一版	外文出版社	中华人民共和国国务院	—	—
《通用规范汉字表》	第一版	语文出版社	教育部国家语言文字工作委员会	—	—
《现代汉语词典》	第七版	商务印书馆	中国社会科学院语言研究所词典编辑室	—	—
《第一批异形词整理表说明》	第一版	语文出版社	异形词研究课题组	—	—
《第一批异形词整理表》	第一版	语文出版社/中华人民共和国教育部、国家语言文字工作委员会	/	—	—
《中国成语大词典》（新一版）	第一版	上海辞书出版社	王涛 等	—	—
《使用语言文字规范指南》	第一版	上海辞书出版社	费锦昌	—	—
《外国地名译名手册》（中型本）	第一版	商务印书馆	中国地名委员会	—	—
《英语姓名译名手册》	第五版	商务印书馆	李学军	—	—
《辞海》	第七版	上海辞书出版社	陈至立	—	—
《中华人民共和国地名大词典》	第一版	商务印书馆	崔乃夫	—	—
《中国人名大辞典》	第一版	商务印书馆	臧励稣	—	—
《中华科学技术大词典》	第一版	商务印书馆	全国科学技术名词审定委员会	—	—

5.5.2 敏感人物检测

内容审核平台对敏感人物的检测，全面识别落马官员、劣迹艺人、失德艺人、涉政正面人物、涉政负面人物、暴恐人物等敏感人物。具体包括国家/地区落马官员、劣迹艺人、失德艺人等。支持检测视频、图片中出现的用户自定义敏感人物。

人脸识别功能是采用深度卷积神经网络（CNN）为核心的人脸识别算法，整合了基于CNN的人脸检测、关键点检测、对齐、特征提取算法作为人脸特征提取流水线。该功能具有人脸检测、人脸定位、人脸特征提取、人脸比对等一整套技术方案，以及高效的数据处理能力。其人脸库具有网络搜索功能，为人脸识别引擎提供高效实时的人脸特征数据。

针对特定人物采用卷积神经网络进行深度学习建立人脸特征模型，可自动识别视频画面中所有人脸的五官及轮廓位置。在保证精度的前提下针对视频优化的接近实时的人脸识别模型，可自动识别视频画面中所有人脸的五官及轮廓位置，返回人物名称、人脸位置信息、人脸表情、视频人脸出现的时间区间。

5.5.3 涉政

识别是否包含涉政内容，包括党和国家领导人、党的政策方针、反动言论、政治事件、外国政要等。

以模拟人类对自然语言的理解，并返回用户所期望的结果。对采集到的文本数据进行自然语言处理，提取特征词、关键词、实体名称并且根据句法逻辑对相关的情感倾向进行初步判断。

根据负面词意、词频、以及负面词出现的位置、标签类别等多维标准，对负面信息的重要性程度进行分类。并通过机器学习的方式强化数据准确性，以及标记的稳定性标签分类。

根据涉政词库、涉政人物关键词等对数据做文本比对分析，识别其中涉政词汇并对用户进行预警。

助力互联网内容责任主体筑牢安全防线。巡检建党百年提法、落马官员、领导人书写错误、领导人外号变体、反动人物等，以及涉及台湾和其他敏感的政治性错误。同时基于海量人脸库和专业审核人员的审核标准，利用深度学习技术，识别正常、漫画、恶搞、负面涉政人物的违规信息，降低违规风险。

5.5.4 暴恐违禁

通过海量暴恐图片库，依托深度学习引擎，支持国旗国徽，恐怖主义，军装，枪支道具，血腥暴乱，儿童邪典等不同类型的图片识别。恐怖主义包括主要恐怖组织标识、台独、港独、藏独、疆独等。血腥暴乱包括血肉模糊、燃烧爆炸、游行集会等场景。

5.5.5 违禁广告

通过OCR识别、垃圾文本识别、广告分类模型，利用强大的文字变体识别能力，上万种主流联系方式（微信、qq、手机号、网址、公众号、百度搜索、微博、广告法合规等）变体特征库。

5.5.6 色情

主要识别是否包含涉黄内容，例如谈论敏感器官、开低俗玩笑、发表性骚扰言论等。

以模拟人类对自然语言的理解，并返回用户所期望的结果。对采集到的文本数据进行自然语言处理，提取特征词、关键词、实体名称并且根据句法逻辑对相关的情感倾向进行初步判断。

根据负面词意、词频、以及负面词出现的位置、标签类别等多维标准，对负面信息的重要性程度进行分类。并通过机器学习的方式强化数据准确性，以及标记的稳定性标签分类。

采用以情感词为文本特征的SVM情感分类方法，对数据进行分析，将色情污秽内容分为色情、性感等多个等级，灵活适应不同应用、场景，利用强大的NLP语义模型和大规模GPU集群和深度学习技术，多角度全方位进行拦截。

5.5.7 特殊标志识别

特殊标志识别包括但不限于以下内容：

- a) 支持公共 LOGO 识别，包括常见的品牌 LOGO、电视台 LOGO、企业 LOGO 等。
- b) 支持自定义 LOGO 识别，支持识别图片、视频中出现的用户自定义 LOGO。

5.5.8 辱骂内容识别

主要识别是否包含辱骂内容，包括脏字脏词、训斥指责、人身攻击等。

对采集到的文本数据进行自然语言处理，提取特征词、关键词、实体名称并且根据句法逻辑对相关的情感倾向进行初步判断。根据负面词意、词频、以及负面词出现的位置、标签类别等多维标准，对负面信息的重要性程度进行分类。并通过机器学习的方式强化数据准确性，以及标记的稳定性标签分类。根据辱骂词库对数据做文本比对分析，识别其中辱骂词汇。

5.5.9 民族问题识别

主要识别是否包含民族问题的内容，包括少数民族政策、制造民族对立、攻击少数民族特等。

对采集到的文本数据进行自然语言处理，提取特征词、关键词、实体名称并且根据句法逻辑对相关的情感倾向进行初步判断。再对用户定制的关键词组进行巡查，根据民族词库对数据做文本比对分析，识别其中涉民族词汇。

5.5.10 宗教问题识别

主要识别是否包含宗教问题的内容，包括传教活动、宗教歧视、宗教迫害等。根据负面词意、词频、以及负面词出现的位置、标签类别等多维标准，对负面信息的重要性程度进行分类。通过数据做文本比对分析，识别其中涉宗教词汇并对用户进行预警。

5.6 业务标签体系

应建立健全与内容审核相关的业务标签体系，搭建与标签体系配套的编码逻辑，并及时更新，如表1样例：

表 2 业务标签体系

标签编码	标签名称	标签编码	标签名称
SH01	涉政	SH010001	政治人物
		SH010002	负面人物
		SH010004	公众人物
		SH010005	涉政场景
		SH010006	国旗
		SH010007	党旗
		SH010008	军旗
		SH010009	徽章
		SH010010	臂章
		SH010011	标志检测
		SH010012	反动组织相关标识
		SH010013	地图相关
		SH010014	公务服饰
		SH010015	公务车辆
		SH010016	钱币检测
		SH010017	落马官员
		SH010018	国家领导人
		SH010019	政府官员
		SH010020	英雄模范
		SH010021	反动人物
		SH010022	敏感人物
		SH010023	党政相关违规
		SH010024	领导人相关违规
		SH010025	封杀艺人

		SH010026	文本涉政
		SH010027	种族主义
		SH010028	宗教
		SH010029	南海诸岛缺失
		SH010030	阿克赛钦缺失
		SH010031	台湾岛缺失
		SH010032	藏南缺失
SH02	涉黄	SH020001	男性裸露
		SH020002	女性裸露
		SH020003	艺术品色情
		SH020004	低俗动作
		SH020005	卡通色情
		SH020006	亲密行为
		SH020007	性用具
		SH020008	敏感部位特写
		SH020009	性器官特写
		SH020010	女性性感
		SH020011	色情广告
		SH020012	色情识别
SH03	血腥暴力	SH030001	血腥场景
		SH030002	尸体
		SH030003	排泄物
		SH030004	游行
		SH030005	爆炸火灾
		SH030006	管制刀具
		SH030007	枪支弹药
		SH030008	大型武器
SH04	违禁	SH040001	赌博
		SH040002	违禁品
		SH040003	不良场景
		SH040004	广告推销
SH05	文本敏感词	SH050003	谩骂
		SH050004	涉政敏感词
		SH050005	涉黄敏感词
		SH050006	涉暴恐敏感词
		SH050007	违禁词
		SH050008	广告推广
		SH050009	特殊语种
SH06	暴恐	SH060001	暴恐人物

		SH060002	暴恐识别
SH07	OCR 识别	SH070001	OCR 敏感词

6 技术对接要求

集约化平台根据电子政务相关标准规范提供标准接口，实现平台的扩展。集约化平台将建立统一的接口规范、统一的接口配置管理、统一的接口日志模块、统一的数据交换模块，同时支持多种系统的接入整合。

充分利用云计算、大数据等相关技术，构建以集约化基础支撑功能为基础，兼顾个性应用扩展，具备符合安全防护要求，具有高可靠、高性能、易扩展特性的集约化基础支撑平台，有力支撑政府网站集约建设与创新发展。

实现对所有共性应用进行统一管理，政府机构新增服务，可通过服务治理进行添加，审核通过后用户可通过服务中心进行服务的申请，可对服务的相关接口进行统一管理，可将系统中数据通过接口方式进行同步，保持数据的一致性。具备对服务的注册、接入、启停、调用等统一配置和管理功能。

6.1 接口规范

6.1.1 URL 结构

URL表示资源地址，应采用RESTful风格，使用小写字母，以斜杠分隔，并且尽量精简参数。

6.1.2 数据格式

服务器支持Request和 Response采用 JSON格式，并在请求头指定Accept字段。

6.1.3 状态码

程序中需要实现相应的http状态码，对于常用的状态码及描述，应遵循HTTP标准。

6.1.4 请求方式

支持常见的网络请求方式：GET、POST等。

6.1.5 错误处理

异常处理应采用统一的返回格式，不同错误报文应有不同状态码、错误信息和错误详情等。

6.1.6 安全性

接口应支持安全认证，固定的加解密要求，比如Oauth2.0。

6.1.7 性能调优

支持缓存技术，利用缓存技术提高接口并发处理能力。

6.2 接口类型

6.2.1 基于 WEB 页面嵌套的接口

基于WEB页面嵌套的接口主要用于集成不同系统的WEB页面。

政府网站系统读取用户身份信息，通过URL 或者表单的形式传递到后台，实现身份的传递。集成时，网站系统将用户名，密码和要请求的URL 信息传递到网站适配器程序，网站适配器程序接受到请求后，解析出用户名和密码信息，用这些信息对用户的有效性进行验证。如果验证成功则转向要请求的URL ，否则转向错误页提示错误信息。

6.2.2 基于 API 调用的应用交互接口

提供给其他应用系统或者系统内部模块进行程序调用的接口类型。按照协商好的API接口规范，交互间的业务系统以及交互间的应用模块调用对方提供的API接口进行程序调用，实现应用的交互。

6.2.3 基于 FTP 的文件传输接口

应采用FTP方式实现文件数据的传输。

注：系统内的数据交换有很大的比例都是以文件的方式进行，使用FTP模式具有可靠性、可扩展性、高性能等特性。
