**技术参数**

**一、项目概况**

互联网违法和不良信息严重污染网络环境，严重危害人民群众利益，特别是高占比的色情类赌博类有害信息社会危害极大，同时也是政治类、暴恐类等有害信息的隐蔽宿主。截止目前，网络有害信息智能化综合治理平台已发现色情赌博网站204.8万个，被劫持网站43480个，境内色情服务器8个，同时还发现政治类有害站群5个共313个网站。这为维护网络意识形态安全和提升公共网络安全水平，以及进一步打击线下违法犯罪团伙提供了路径和支撑。实践证明，陕西省委网信办推动的网络有害信息智能化治理体现了党中央国务院和陕西省委省政府关于维护网络空间清朗、捍卫网络意识形态安全的根本要求，也是“以人民为中心”执政理念的生动实践。

为进一步推动网络有害信息智能化治理工作,为彻底整治网络有害信息（特别是来自境外、针对境内的违法和不良信息）提供持续性、稳定性、对抗性支撑，现需对网络有害信息智能化综合治理平台的功能进一步扩充与增强，提升主动发现和治理政治类有害信息智能化水平，切实保障意识形态安全；扩大和提升色情赌博类有害信息检测覆盖面，为线下打击提供精准线索支撑；对我省互联网访问态势进行研究分析，构建我省网民互联网访问画像，实现科学管网治网；加强有害网站封堵治理管理工作，确保精准封堵，及时处理用户申诉，指导互联网企业提升安全防范意识和具体整改等工作。

**二、项目需求**

**1.政治类有害信息样本库**

通过人工分析平台积累的网站文本数据以及已通报的政治类有害网站，手动收集政治类有害信息，建立政治类有害信息样本库，并可对政治类有害样本进行有效管理。

**2.政治类有害样本分析**

通过自然语言处理技术对政治类有害信息样本库进行分析，根据词频分布提取网主页、文章标题、文章内容的关键词，为人工分析提供帮助，并可通过词云对提取出的关键词进行展示。

**3.政治类有害信息检测模型**

通过深度神经网络建立政治类有害信息的自然语言检测模型，利用政治类有害信息样本库进行训练，实现政治类有害信息的智能检测，要求模型可对支持网站主页文本检测、文章标题检测、文章内容检测。

**4.域名筛选模型**

构建政治类网站域名检测模型，利用我省用户每天上网域名访问数据，从海量域名中筛选出疑似政治有害网站进行检测，提高平台检测效率。

**5.政治类有害网站检测**

利用域名筛选模型提取出待检测的目标网站，访问目标网站获取网站主页内容，通过政治类有害信息检测模型对主页内容进行严判，并采集网站文本、主页截图、域名的解析、域名注册等信息。

**6.政治类有害网站证据采集**

对发现的政治类有害网站，解析网站主页内容，通过政治类有害信息检测模型发现政治有害文章标题或视频标题，并获取文章全部内容，经模型研判后，形成证据信息并进行保存，以便人工研判。

**7、政治类有害信息人工研判管理**

展示发现的政治类有害网站及证据信息，工作人员通过查看网站截图、网站文本内容、证据信息，进行人工研判，对人工研判确认的政治有害网站，将相关信息存储到政治类有害信息样本库中。

**8.政治类有害信息人工检索与分析**

工作人员根据已掌握的政治有害言论，定义基于文本特征的匹配规则，平台根据定义的规则对已积累的网站数据进行分析，发现疑似政治类有害网站后，交由人工分析确认。

**9.政治类有害网站封堵管理**

对已确认的政治有害网站域名进行封堵管理，详细记录每次发布的封堵名单。定期检测网站活动情况和内容，网站不再活动后解封。

**10.政治类有害信息统计与分析**

根据年度、月度对政治类有害网站发现情况进行统计，分析政治类有害网站分布情况、域名注册情况、域名解析情况、CDN使用情况等，掌握分布特征。通过用户每天上网域名访问数据，对每个政治类网站进行回溯分析和监测，掌握每个政治类有害网站的访问态势。

**11.支持多顶级域的色情赌博类网站检测**

在一期基础上进一步优化检测模型，扩大检测范围，支持对色情赌博类网站存在较多的vip、xyz、top等多个不同顶级域的检测。

**12.支持实际流量数据的****色情赌博类网站检测**

利用域名、解析IP等基本信息，构建色情赌博类网站域名检测模型，快速对用户访问日志进行分析与检测，能高效、准确筛选出疑似色情赌博类有害网站使用的域名。

**13.色赌网站图谱关系检测**

对已发现的色情赌博类网站进行全面检测与分析，对与之关联的网站及URL链接进行检测，拓展现有色情赌博类有害信息检测方式，进一步提高检测效率与检测范围。

**14.专用色情赌博DNS服务器检测**

利用平台积累的数据，发现为色情赌博类网站提供域名解析服务的专用DNS服务器，对专用DNS服务器部署情况进行分析，发现部署在境内的专用DNS服务器。

**15.专用色情赌博图片服务器检测**

利用平台积累的数据，发现为色情赌博类网站提供图片存储服务的专用图片服务器，对专用图片服务器部署情况进行分析，发现部署在境内的专用图片服务器。

**16.专用色情赌博视频服务器检测**

模拟访问色情网站中的视频链接，并监测浏览器的网络行为，发现提供服务的视频服务器，同时获取视频服务器的解析信息、注册信息等。

**17.色情赌博类APP下载及管理**

对色情赌博类网站中包含的APP进行下载，通过聚类算法对已下载的APP进行聚类分析，将代码相似的APP归集到统一家族，以便定期人工对推广排名高的APP进行人工分析，对色情赌博类APP进行治理。

**18.会话劫持检测**

平台在检测过程中，对每个网站的访问过程进行监测，确认网站在访问过程中，是否存在会话劫持，也可对域名进行主动探测，分析在访问过程中是否存在会话劫持行为。对存在会话劫持的网站进行分析，推断劫持发生的网络位置。

**19.封堵网站的视频取证**

色情赌博类有害网站在访问过程中往往存在访问跳转的现象，要求对封堵网站通过视频的方式详细记录网站访问过程，形成完整的证据信息。

**20.色情赌博类网站封堵管理**

根据封堵工作流程，对色情赌博类有害网站封堵名单管理，详细记录当前正在封堵域名及历史封堵过的域名，并能够利用用户域名访问数据监测封堵治理效果。

**21.企业申诉管理**

对企业申诉进行管理，详细记录企业申诉内容，对企业信息、企业网站整改报告、网站解封等进行管理，实现企业申诉全生命周期管理。

**22.提取违法网站线索**

色情赌博类有害信息与网络诈骗有密切关联，利用已掌握的诈骗线索的关键字、网站文本、网站截图等信息，对已积累的数据进行分析，发现相似的诈骗网站。

**23.专项赌博封堵策略**

根据网站文本、主页截图建立检测模型，对已发现的赌博网站进行类别细分，如：体育竞技类赌博、彩票类等，并根据类别快速提取网站域名。

**24.建立网站分类库**

周期性统计用户每天上网域名访问数据，提取省内用户访问排名前1万的网站，根据各个网站内容，建立知名网站分类库，如：综合门户类、视频类等。建立我省政府网站、知名媒体网站等信息库。

**25.全省域名访问态势**

利用每天上网域名访问数据，分析我省各个运营商的网络活动情况、各个顶级域中的域名活动情况、互联网资源分布情况、省内网民访问资源分布情况等，掌握我省互联网访问态势。

**26.省内网民访问特征**

利用网站分类库，分析我省网民互联网活动特征、国内重要媒体平台的访问情况，建立我省互联网访问画像。

**27.省内重要站点访问态势与监测**

分析省内网民对省内政府、省内重要媒体等网站的访问态势，掌握各个网站的每天访问情况，对访问异常的网站实时报警，以便工作人员及时分析、发现风险，如：潜在舆情事件等。

**28.各类非法信息访问态势**

监测省内网民对政治、色情、赌博类等违法和不良信息访问态势，主要包括各类有害信息的每天请求总次数、每天活动网站数量、每天有效封堵的网站数量及封堵次数、有害类网站访问排名等。

**三、技术要求**

1.总体要求

系统采用B/S体系架构，具有平台无关性，兼容360浏览器、微软Edge、Chrome等主流浏览器。软件开发技术要求如下：

1.1规范性：平台的设计和开发均需严格遵循国家互联网信息办公室及陕西省互联网信息办公室制定的各项规范和政策。

1.2通用性：平台在框架设计上，保证软件基础设施有足够的通用性，以便实现计算资源的共享，为日后其它业务应用系统的开发、部署奠定基础。

1.3稳定性和可靠性：软件设计必须保证系统的稳定性和可靠性，支持负载均衡机制。

1.4可扩展性：平台的功能设计、数据库设计等方面必须具有良好的扩展性。

1.5易用性：用户界面设计友好、美观，操作简单、实用，重易用性。

1.7 安全性：确保系统较高安全性，检测数据不被泄密。

2.准确率与漏报率要求

为有效对色情、赌博、政治类有害网站进行治理，各类智能检测算法要求：准确率不低于90%、漏报率不高于10%。

3.性能及稳定性需求

要求科学的软件设计架构，以保证系统具有高性能、高检测效率及稳定性，具体要求如下：

1）平台每天可完成研判域名数量不低于10万个；

2）系统正式运行后，要求7\*24小时稳定运行，年可用率≥95%；

3）平台支持并发用户数不低于50个；

4）WEB应用系统响应时间低于3秒。

4.可扩展性需求

采用组件化设计思想，平台基础架构与业务逻辑分离，在不改变架构的前提下，可实现系统功能的扩展，且不影响原有功能正常运行，平台能够根据检测规模的需要，动态扩容。

5.数据要求

1）承建单位能够提供vip、xyz、top顶级域的相关数据资源，并每天可及时更新，保证每天可及时掌握这些顶级域下的新注册域名、过期域名等数据信息。

2）人工收集政治类有害样本数量不低于1万条；

3）知名网站分类库至少包含访问排名前1万的网站。

6.模型周期更新要求

为保证平台的检测有效性，能够周期性对各类智能检测算法进行训练及修正，要求每半年重新训练优化一次。

7.安全性要求

根据《中华人民共和国网络安全法》，本系统上线前需通过第三方安全检测。系统上线前必须经过严格的漏洞扫描，不允许存在SQL注入、session劫持、跨站脚本攻击等漏洞，降低被攻击的风险，另外，涉及保密数据、用户信息等敏感数据的传输必须加密。

**四、技术培训**

供应商需要提供软件操作培训,确保工作人员能够正确使用平台，应在响应文件中提交培训计划书，提交详细的培训计划、培训内容，列出详细的课程安排及时间表，并提交采购人审核。