

人工智能多媒体内容审核技术

一、系统运行环境

- 1、服务器操作系统： Ubuntu 18.04 Server 版本系统
- 2、服务器 CPU 为志强四核及以上，内存 32GB，硬盘 1TB 以上 SSD 固态硬盘
- 3、服务器 GPU 为英伟达 RTX1080Ti/RTX2080Ti 或更高级别

二、内容审核分类

基于人工智能技术，利用大数据训练模型进行自动化的内容审核分类：

- 1、图片分类：正常图片、性感图片、色情图片
- 2、文字分类：涉政、涉黄、涉赌、涉毒、涉恐

三、内容审核性能

- 1、文本文字分类：每秒 1000 字的文本自动化审核
- 2、图片涉黄分类：每秒 10~100 张的图片自动化审查
- 3、图片转文字：每图片 0.1~10 秒（图片分辨率和文本数量相关）
- 4、APP 内容测试支持 7x24 方式监控，需要编写脚本协助

四、内容审核技术

- 1、支持文本内容涉政、涉黄、涉赌、涉毒、涉恐等违规内容的自动化审核
- 2、支持图片转文字，并对文字进行涉黄、涉政等违规内容的自动化审核
- 3、支持图片内容涉黄违规内容的自动化审核

五、内容检测流程

（一）APP 内容审核流程

- 1、自动化安装 APP 到测试终端，采取定义的脚本自动遍历 APP 功能模块，或者采取随机方式自动遍历 APP 功能模块
- 2、自动化运行 APP 期间，采集 APP 在真实手机端的图像信息，进行文字内容、图像内容安全的审核比对
- 3、对于审核内容比较敏感的信息，自动上报到平台，提交人工审核确认

（二）文本类内容审核流程

- 1、建立自动化文本内容审核任务，指定任务名称，选择目标测试文本文件，文本格式为 Text
- 2、任务提交生成后，自动化进行目标测试文本文件的内容审核检测，对于匹配涉黄、涉政等违规内容的文本信息做自动化标记

（三）图片类内容审核流程

- 1、建立自动化图片内容审核任务，指定任务名称，选择目标测试图片文件，文本格式为 png、jpg 等
- 2、任务提交生成后，自动化进行目标测试图片文件的内容审核检测，自动进行涉黄分类识别和标记；自动提取图片中的文字信息，对于匹配涉黄、涉政等违规内容的文字信息做自动化标记。