

QQ 频道内容安全审查实施细则

一、执行机构

灵域委托旗下子公司以及阿里云对 Agatha 纯净生存 QQ 频道内的信息元数据进行内容安全审查。

二、审查细则

所有审查项包括文字和图片。

1. 基线审查

基线审查是保障内容符合最低安全要求的检查，主要审查依据包括相关法律法规，保证频道内出现的内容均合法。

该项审查由阿里云执行。

涉政

敏感政治事件、社会舆论事件、历史虚无、反动分裂、反华对立、反动人物、非法团体、特定时间节点上线的涉政专项保障监测机制

色情

涉黄产业、淫秽影音文资源、性暗示、低俗内容

暴恐

极端组织、武器弹药、涉恐音像内容

违禁

涉黄赌毒、犯罪活动、突破跨境数据安全网关、违禁书籍、电信诈骗

宗教

宗教极端主义、恐怖主义、涉基础教育宗教问题

不良

种族主义、封建迷信、网暴言论

以上项目均可机审机评。

2. 内容导向审查

内容导向审查是保障频道内风向正确、内容积极向上、信息真实有效的审查机制。该项审查由灵域执行。

重点审查项

- ① 有意展示色情擦边内容，含特定色情特征、伪色情、软色情、个人癖好，特别是可能符合此列表第⑧项的内容
- ② 有意煽动暴力、怂恿他人参与或向他人展示可能造成人身伤害、死亡的内容
- ③ 有意对个人或群体进行嘲讽、发布仇恨言论、挑拨他人矛盾，特别是可能符合此列表第④项的内容
- ④ 有意鼓动其他用户发动对玩家管理、运行组、灵域及其员工、子公司员工的攻击，以及网络攻击，特别是网络暴力的内容
- ⑤ 引流至其他应用程序，特别是游戏类的内容
- ⑥ 发布易造成恶心以及其他引人不适的内容，特别是排泄物的内容
- ⑦ 发布极端民族主义观点、种族主义观点，特别是在基线检查被拦截时的内容
- ⑧ 使用冒犯性语言，特别是语言性骚扰的内容

- ⑨ 有意规避各项审查机制，使用各种手法绕过审查的内容
- ⑩ 其他法律法规、社会生活所不倡导的内容

以上项目均可机审机评。

3. 建模审查

建模审查是指对玩家个性特征进行分析的审查。
该项审查由灵域执行。

建模维度

- ① 外倾性
- ② 随和性
- ③ 责任性
- ④ 情绪稳定性
- ⑤ 创新性

以上项目人工修正，不作为任何决策判断依据。

三、决策处理

风控系统由上述三层审查确认特定用户存在违规行为，可以自行作出撤回消息、撤回内容、禁言的决策，但不能作出封禁用户的决策。由内容安全审查检出需进行封禁操作的用户，需要投票或由玩家管理、运行组二次确认。