

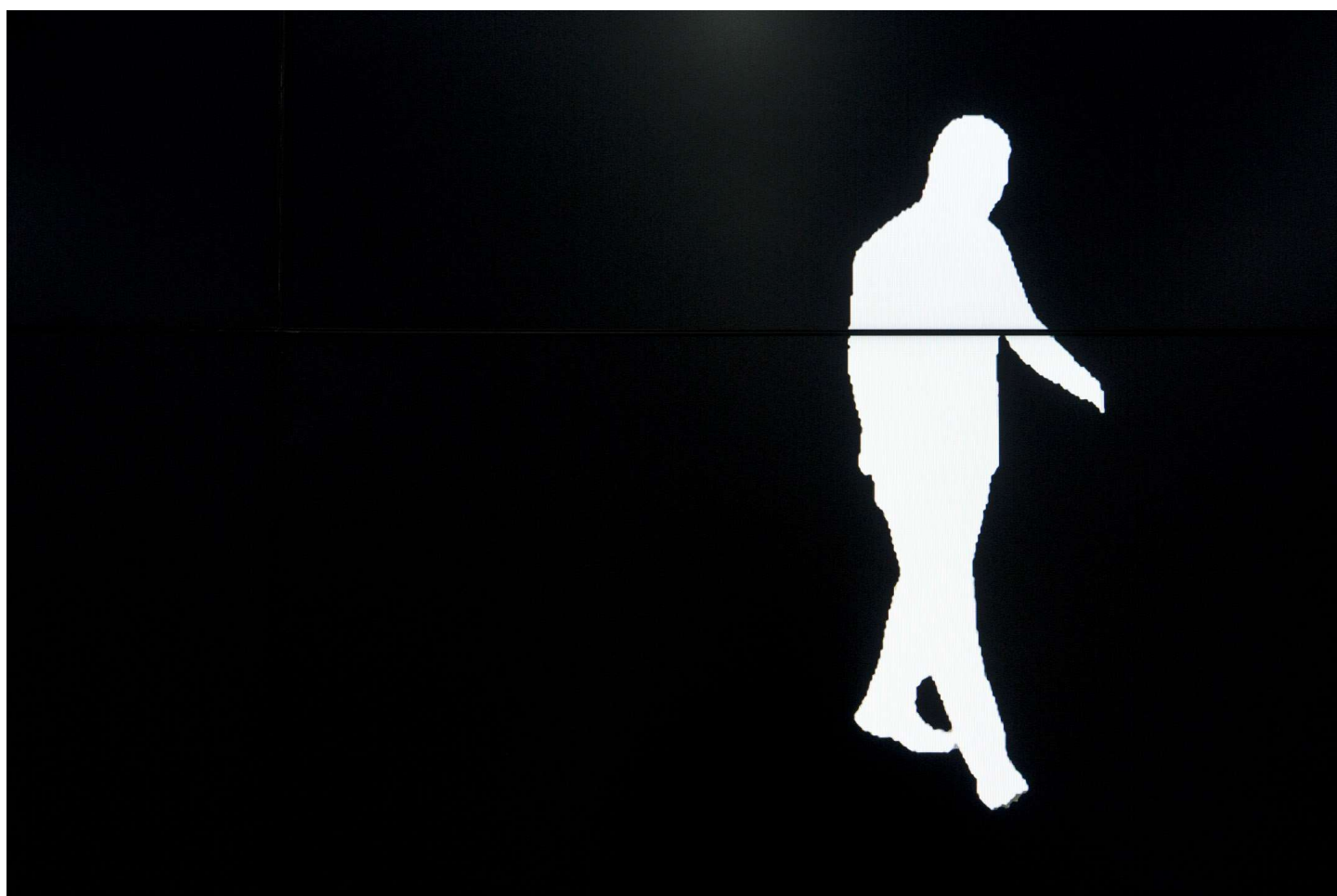


深度

网络审查员自述：以后决定互联网平台发展的，是审核团队

我每天看1200多篇文章。我不关心政治。如果以后大陆像台湾一样可以反思“二二八”，我们再来讨论审查合不合理。

特约撰稿人 路屿 发自香港 | 2019-03-14



中国大陆有8.29亿网民，他们人均每天上网四小时，生产、接收巨大数量的信息。其中的绝大多数内容，都是经过审查的。没有数据统计中国目前有多少人在从事审查工作。图为一间科技公司的步态识别软件演示，可能有助于中国公安部门查明嫌疑人。摄：Imagine China

2月，中国网络视听节目服务协会发文，要求视频网站对“弹幕”进行“先审后播”，网络审查的“版图”再一次“扩张”。中国大陆有8.29亿网民，他们人均每天上网四小时，生产、接收巨大数量的信息。其中的绝大多数内容，都是经过审查的。没有数据统计中国目前有多少人在从事审查工作。随着相关审查政策不断出台，各大互联网平台都在数以千计地扩充审查团队。在拥有逾2亿日活跃用户的今日头条，审查队伍已达万人。

端传媒深入采访了大陆某互联网平台的一名审查员，希望通过ta的自述，展现审查员的工作内容、流程，以及他们是如何看待这份工作的。

培训时讲敏感性事件，大家会发出“哇”的声音，“原来还有这么一段历史”

两三年前，有朋友推荐我来这家互联网公司，面试过了，就一直做到现在。我以前就知道有内容审查这类岗位，但我想像的是关于色情低俗方面的，没往政治方面考虑。

公司的招聘需求以大学本科为主，实在招不到会放开招一部分专科。党员肯定是会优先录用的，这个会写进（招聘公告）去。我们平均年龄在26岁左右。

我们面试很简单，会问他们能否承受色情、暴力的内容，是否能承受夜班，基本不用怎么考虑他们的意识形态，因为从小接受的教育就是这样，思维模式已经形成了。

以前面试会问敏感的内容，直接问你知不知道（某个政治敏感事件），但慢慢发现这样是有风险的。比如，有人参与面试，被问到郭文贵（编注：流亡海外的中国地产大亨，2017年开始持续在网络爆料，批评中国政府高层腐败行为，被中国政府通缉。目前居于美国，已申请政治庇护），一脸迷茫。好多人是不知道这些敏感内容，如果你不知道的话，回去会查这些内容。如果这个人通过了（面试）还好说，没通过的怎么办？相当于从我们这里把那些敏感内容都泄漏出去了。

后来我们就改成考验对时政和热点新闻的解读，比如你怎么看待假疫苗事件。基本上你只要能很好地表达出自己的观点就可以，没有具体的套路。来面试不会批评政府的，即使心里有也不会说。

入职会经历两轮培训。第一轮，公司会对新入职的进行大范围集中培训，给他们讲一些比较敏感的问题。大家都是不知道的，他们是真的不知道。上了大学的都不知道六四。



1989年6月4日，天安门事件。摄：Bettmann / Contributor/Getty Images

培训的内容主要是政治类，先告诉你个大概。细分会有国内的和国际政治，以及恐怖组织、非法活动等。非法活动主要是邪教组织，法轮功、3K党之类。会专门培训西藏、新疆、台湾这几个敏感区域。

低俗类按每个人的三观也能辨别出来，但政治类你以前就是没有接触过。相当于历史课，前前后后全都给你讲了。从它的背景、开始、发生过程、结束，以及它的后续影响。会讲现行历史书以外的东西，像孔庆东、王丹、天安门母亲、天安门四君子这些人物，会详细到每个人做的事情。像吾尔开希和赵紫阳的内容也会直接放出来。这些翻墙都可以看到嘛，我们也会拿过来看看了解一下。以前六四的背景，东欧剧变、苏联解体也是一笔而过。大概了解了再告诉你，如果遇到了应该怎么处理。当时我入职的时候，培训时不会评价六四谁对谁过，只是说这是一个历史性事件。

我们有一个专门的培训部门。对培训部门的学历和背景要求没有很高。一个培训部门有一个人知道这些内容就可以了，再传授给别人。

他们第一次培训的时候，如果你经过门口，就会听到“哇”的声音，反应还是比较强烈的：“原来还有这么一段历史”，“原来还有这么一些故事”。

但是过了一段时间，就没什么感觉了。只要你确实没有侵犯到个体的利益，（大家）对于政治基本上都是漠不关心的。很少人会在培训中，价值观受到动摇，更不会为此离职。即使动摇，在工作中还是会该怎么做就怎么做。我们在工作中也不会聊这些事，感觉没什么意思。只谈风月，不谈政治，是不是？

比较大反应的还是暴力血腥类的，有的人会因此离职。

我记得之前遇到一个斩手的视频，我看到后，立刻把电脑合上了，找人处理掉。我希望我从来没有看到过。你会好几天都很难受。公司也有专门的心理辅导部门。

刚开始培训时还没有签正式合同。有的人承受不了，也就走了。你想想“新疆七五”事件（编注：2009年7月5日，新疆乌鲁木齐市爆发民族冲突。据官方统计，大规模的骚乱和暴力活动造成近200人死亡，逾千人受伤。此后中国政府不断加强对新疆的管控。）和“西藏三一二”事件（编注：2008年3月初，连续数日发生藏人僧侣绝食抗议事件，僧侣被政府抓走后，藏人于3月14日进行大规模游行，要求释放被捕僧人，事态演化为大规模骚乱，而后中国军警镇压，致多人死亡，数百人被捕）都是很血腥暴力的，尤其“七五”的现场画面，全是血、全是死人。你感觉不合适的话，随时可以走人，也会给你一定的补偿。



2009年7月，新疆乌鲁木齐市爆发民族冲突。摄：Guang Niu/Getty Images

新的审查员犯错会比较多。但在他们来之前，我们会有一个考试系统，模拟性地审核。当正确指标达到90%以上，那就可以正式上线了。如果正确率非常低就代表不适合这个工作了。公司培训很简单，业务线培训比较难。犯错最多的还是色情和低俗类。

考核的东西是我们遇到过的案例汇总。比如在短视频中，有人把达赖喇嘛的画像挂在家里的客厅，如果（这类）背景出现，不管这个视频内容多么积极向上，也要处理掉。题目中会有人脸识别，放上敏感人物的照片让你选择；还有我说的隐蔽的背景图片形式，问你审核时是否能通过；基本上就是单选和多选题，考察你对于敏感人物和事件的了解程度。

通过考试后，再根据内容产品的类型，分到各个独立的业务线，然后会有第二轮培训，我们称之为入组培训。第二轮培训和第一轮最大的不同是，详细到你会遇到的每个情况，都告诉你该怎么操作。

机器先审人再审，每天要看1200多篇文章

审查的具体流程，基本上是机器先过一遍，没问题的就放到人工审查，审完就会放出了。放出后如果浏览量达到一定的值，就会进入到另一个模型里面再审核，如果浏览量又达到一定高度，就会进入到代表热门的标签里，然后会再看一遍，如果没有问题就会再放出。每小时能审150篇左右。

会出现一开始没发现、后面才发现有问题的情况。人和机器都会出现，机器出现的概率大一点。基本上是色情类的问题。很隐晦的色情类你可能一下子看不懂，就放行了，第二个人可能还是没看到，最终展示量达到一定高度后返回了，第三个人看出来了，这个时候相当于前两轮都放出去，而第三轮却拦截下来。这样的情况肯定会有。

政治类的内容还没出现过（类似情况），政治类太敏感了。算法基本上对低俗的识别能力没有政治类的识别能力高。政治类的放出，更多是预警没有跟上。比如前段时间的《还愿》，在2月23日中午12点之前是全部放出的，也非常火，但是11点47分的时候，被举报了，跟着那条新闻火了后，才做全都下架的处理，整个大陆都是这样的。这个和事件突发性有关。



今天所有的门户网站会处理掉关于赤烛的所有内容。因为通过《还愿》这个游戏已经给这家公司做了定性，这是台独的公司。网上图片

基本上网信办每天都有指令。只要网信办发出的指令，都是严重的。指令基本上是根据时事和政治走。比如中美贸易谈判里面，达成什么样的协议；还有金正恩去越南和特朗普会谈，关于金正恩横穿中国产生的一些不好的影响、网友的评论，全都要处理掉。比如导致交通堵塞，有的评论说“劳民伤财”之类的，这种就不能出现。所有门户网站都有风险控制部门、GR（Government Relationship）部门，都会有专门的处理。公司层面自己操作的话，即使不删除也会做一些处理，比如只能订阅可见，只能（用户）自见。

据我所知，自从我来公司，没人在敏感话题上出现过失误，这个很难出现。

在敏感时期来临之前，我们会把以往的案例稍微梳理一遍，根据经验来制定一个规则。以六月为例，会把所有可预测到的内容、可能出现的人物梳理一遍。他们就那些人物嘛，每年来来回回就那几个人。也会出现暗语，把这几个字连起来，或倒过来写，我们发现了就会解决掉。这样的情况，就需要一个审核员的敏感意识了吧。每天都需要一篇篇文章去看，每天一个人累死累活能看1200多篇文章，差额在20%上下浮动。像六月比较敏感，进新人就会比较少了，没有那么多时间培训。

预测不到的敏感点，就需要主动以最快速度去解决。你看开两会了，肯定舆论会更加严格一些，我们会针对两会进行培训。首先是两会的场景，比如人民大会堂啊、他们开会的场景是不允许出现在自媒体上的。比如一个人大代表把开会场景拍下来，然后上传到视频网站，是不会过（审）的。还有领导人长什么样也会重新培训一遍。如果在家里拍个照，后面有个领导人画像也是不行的。因为你很难判断自媒体会发布什么内容，政治风险性很大。你把“发表”写成“发飙”（编注：腾讯网曾将习近平“发表重要讲话”错发成“发飙重要讲话”），影响都那么大，所以宁愿全都处理掉。其实其他时期标准也一样，只不过两会期间如果出现问题，可能影响更大。

审核的时候会有质检流程。1000条内容我们会抽出10%盲审，同时两个人去审。如果两个人不一样了，就会到第三轮，到质检队列。进入到质检队列，又会有一个人去审，所以会有三个人去审一条。基本上第三个人审，就会对这两个人进行判定谁对谁错。当然每个人都可能有错误的时候。犯错只有非常严重的时候，可能会被开除，其他时候我们还是比较

人性的，（犯错了）可能在绩效上降低一个档次。不可接受的，有政治错误和严重的色情错误。

我们有夜班嘛，晚上值班的时候一打瞌睡，一敲空格，过去了也没看到。业内某家的员工，夜班打迷糊，出现一个很严重的色情类错误，导致App下架了。触犯到应用商店的规则，或是被当地政府监测到，就会做下架处理。下架了一个星期，你想想损失有多少。



2017年11月28日，郭文贵在美国寻求庇护。摄：Timothy A. Clary/AFP/Getty Images

一个公司要做得好，就得既能保住安全，又能保住质量

对一些案例，业内的态度是完全一致的。制造《还愿》游戏的赤烛公司，它做过好多游戏。今天所有的门户网站会处理掉关于赤烛的所有内容。因为通过《还愿》这个游戏已经给这家公司做了定性，这是台独的公司。再比如，金马奖颁奖时，出现的涉台独言论。我们会一律全部下架处理。别的门户网站也一样，同样看不到。

对一个新的审查员来说，比较难掌握的还是低俗和色情类。在我们部门，离职的人一般是不能接受色情方面的内容。有个人的离职申请理由是，色情内容会影响他。因为这个度实在不好拿捏，什么样是低俗，什么样是色情，什么样又叫艺术品，刚来的时候不太好掌握。有的人看到大卫，就定性为色情。

以前审查是安全思维——仅保安全，什么都不要。现在慢慢转换为质量思维，不仅要保安全，还要控制好质量。不可能因为大卫裸露，你就处理掉、下架了，它是一个高雅的艺术形式。

一个公司要真正做得好的话，关于内容和体量，其实你需要既能保住安全，又能保住质量，比如艺术品、预防爱滋的、红丝带、预防乳腺疾病的等等，用各种艺术形式表现的，这样好的内容是有利于整个发展，是能留下来的。如果全都误伤了，对于体量是很不好的。

政治内容没有空间，它就是那样的。像李洪志是法轮功的，大家都知道他做什么的，发出来影响也不好，我们肯定不会替李洪志说话。这个大家都明白。这样是属于底线性质，不会有多大的分歧点。

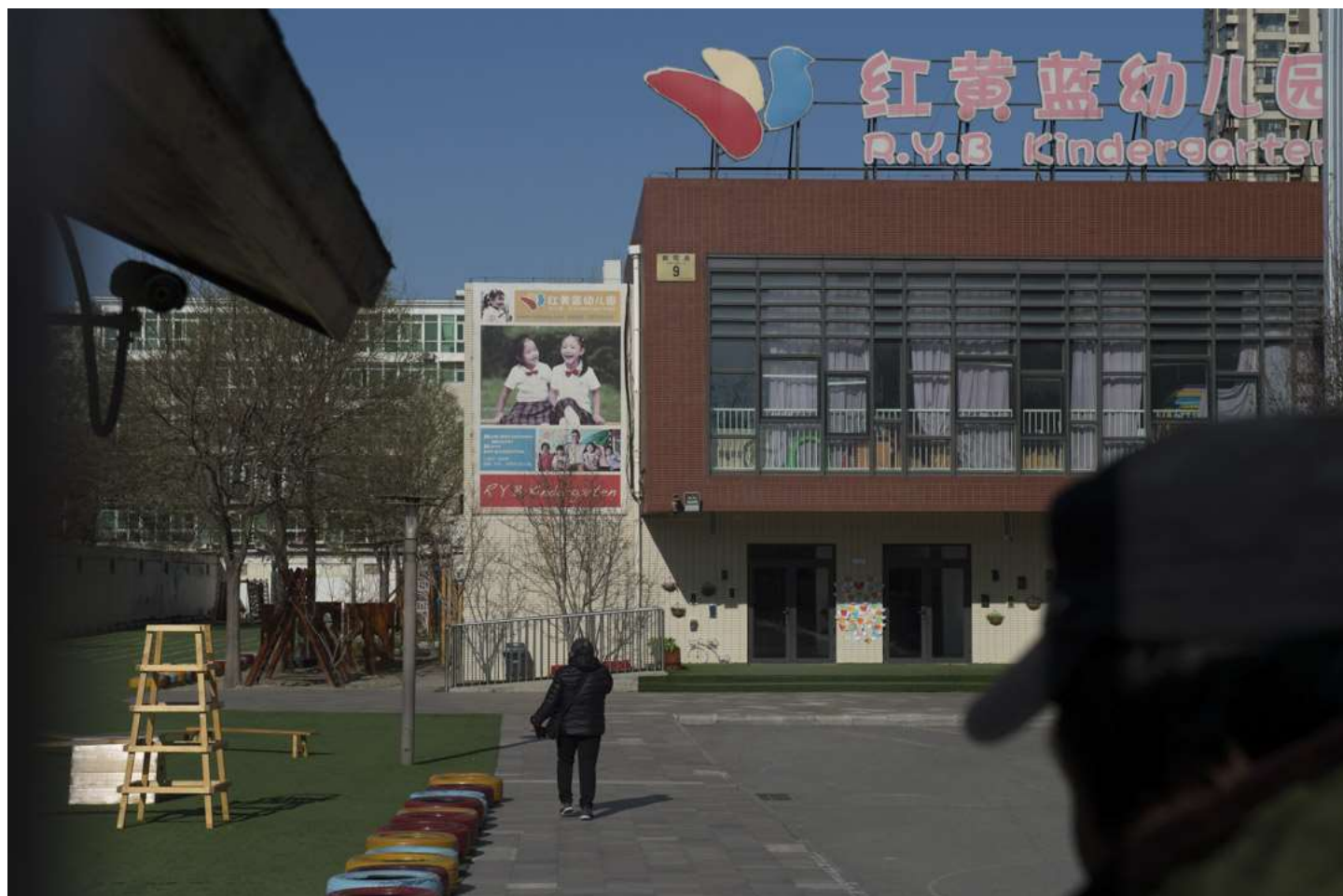


2001年5月9日，法轮功学员在香港湾仔入境事务大楼外练功。摄：Peter Parks/AFP/Getty Images

社会事件的话，像“红黄蓝”（编注：2017年11月，北京红黄蓝新天地幼儿园被指涉嫌性侵和虐童事件，引起社会广泛关注，2018年一名涉案教师被判刑。红黄蓝集团旗下幼儿园曾多次被曝虐童事件），基本上我们往外放的都是正规新闻媒体，自媒体发的和正规新闻媒体差距很大，可能产生虚假消息，很难保证真实性和可靠。所以重大社会突发类事件，是有特定的白名单的，日常接触到的大陆官方媒体都在白名单上，属于正确的舆论引导吧。

整个公司离职率很高，差不多3成这样。因为可替代性太强。我看有报道说，这个工作是信息行业的富士康。其实我也挺认同的。这个工作，你一时半会也做不来什么成绩的话，看不出来能创造什么价值，没有直接性的体现，可能就会离职吧。

我们审查员是分等级的，一二级仅需要知道规则就行了，单纯是一个支持性的岗位，是配合运营的，就和客服是一个性质。你培养安全思维很容易，周期性短，但再往上的话，质量思维就比较难了，你的水平可能要提高很多。我们每年都有升级的机会，但只有很少的人可以往上走，5-10%吧。



2017年11月，北京红黄蓝新天地幼儿园虐童事件，引起社会广泛关注。摄：Nicolas Asfour/AFP/Getty Images

“如果以后大陆像台湾一样可以反思‘二二八’，我们再来讨论审查合不合理”

其实这份工作的价值，说大也挺大，说小也小。比如你把对未成年不好的信息，通过你来解决了，色情低俗的、政治敏感的，这些不好的事情被你解决掉了，你就相当于一个安全卫士吧，保护了更多的未成年。我们曾经在审核中发现有人直播自杀，然后通过我们和公安部门的努力解救了一条生命，是不是很有意义？但这个工作的现实收益很低。

公司每逢重大活动、节假日都会给大家强调重要性，也会讲一些情怀，当然我知道情怀不能当饭吃。领导会把你说的很高大上，会给你说这是非常重要的工作，没有内容审查就没有公司。

可能以后决定一个互联网内容公司发展好坏的，更多的是审核团队的构建，你的审核团队决定了这款产品是否存在。如果审核团队出现问题，就是致命的问题。未来发展应该会是这样层面吧，不然审核团队不会越建越大。

有的时候你会有那种感觉，对内容的生杀大权掌握在我们手里，我让它放就放，不让它放就不放。但其实我们只是根据规则去执行。

审核人员制定规则是很少的。我提出这条规则，会有多方评估，包括风险评控的、运营人员、产品经理的评估.....不可能只有一个人。审核人员是属于这条链条的最底端。大的规则从整体市场确立的时候就确立下来了，根据互联网行业还有国内的标准。

审核的监管可能会越来越严格，会让自媒体这些越来越正规，像各个公司也会主动拥抱这些（审查）。这种严格的管控会逐渐成为一种现实。管控也是应该的，不管是大陆还是其他地区都是一样的，但不代表所有的管控都是合理适当的。



2017年，一位“228受害者”的儿子，于照片中指出当年被处决的人。摄：Sam Yeh/AFP/Getty Images

管控一些突发事件的报道，可能是为了压制舆论，维护稳定。虽然会造成误伤，但也是必须的。西方的言论自由也是在一定的范围内，不可能什么都能说，什么都能放。大陆地区比较敏感的事件，我也不知道到底是真是假，是对是错，站在现在的角度来说可能是合理的，如果以后大陆像台湾一样可以反思“二二八”，那我们可以再来讨论审查合不合理。

我没有后悔的时候，因为漏放比误删严重多了。你漏放一个可能会影响安全，你误删一个仅仅是用户体验。还有一点，这仅仅是一项工作，没有太多个人感情。

对我来说，就是当作一个历史知道就好了，也不会和别人说，我也不会和小孩说。那些东西到他长大的时候可能就会知道了。你看，那些历史事件，我爸妈也没和我提过，你爸妈也没和你提过，大家都没提过。让小孩活在美好当中会更好一些。



热门头条

1. 没有伤口的职灾：那个替乘客脱内裤的长荣空姐，以及她的“制服病”
2. 《还愿》情绪风暴之后，被吞噬的中国游戏人
3. 重磅调查：被强制绝育的淳子的一生
4. “不是日本人”的日本人：国族神话下的北海道爱努族
5. 网络审查员自述：以后决定互联网平台发展的，是审核团队
6. 中国女工生育之痛：被强制流产后，计生人员又把埋掉的孩子挖出来
7. 白信：新疆的“汉化”，还是内地的“疆化”？——新疆纪行之三
8. 我是台湾人，在蒙古国教中文
9. “正妹、高分、老残穷”：这群台湾高中生对放榜新闻说不
10. 白信：冻结的堡垒城市，与消失的人——新疆纪行之一

编辑推荐

1. 一厕一世界：香港公厕折射的管治思维
2. 再访 Ian Johnson：宗教复兴遭遇收紧治理，中国要打一场“灵魂之战”吗？
3. 网络审查员自述：以后决定互联网平台发展的，是审核团队
4. 阿尔及利亚之春：街头抗议浪潮，能否成功逼退掌权二十年的政治强人？
5. 《滚滚红尘》数位修复版：三毛离张爱玲有多远？
6. 量多写不完、循环欠债、形式老旧：台湾小学生成为家庭作业的奴隶？
7. 羊毛党是个什么党？他们立志“吃干榨尽这人间”
8. 田禾：如何科学地反驳诺奖得主沃森的“种族决定智商论”

9. 机械存在主义：当人造人开始思索“人生”的意义

10. 她是菲佣，也是摄影师：“我希望告诉人们香港有多美丽”

延伸阅读

新技术教你如何反网络审查、恢复被删帖文、保护用户隐私

全面审查时代：中国媒体人正在经历什么？

这是一篇中国当代新闻传媒从业人员的口述史，在全面审查时代下，他们无一例外地经历了越来越不自由的从业状态。

审查面纱下的大国形象：2018年被微信删掉的十条新闻

2018年，中国网络审查最敏感最无法触碰的，已经不仅仅在“政治敏感”上——中国网络审查的目光开始转向与“大国形象”相关的内容。

《还愿》风波，会不会是压垮Steam中国版的最后一根稻草？

《还愿》仅是“Steam中国版”正式降临之前的一桩惊悚章节，真正令人焦急等待的还在后头——当电脑里的游戏世界也分“墙内墙外”，或许真正的恐惧才要降临。

《还愿》情绪风暴之后，被吞噬的中国游戏人

《还愿》在中国大陆有多受欢迎，事件反转的时候，玩家就有多难接受。这场争议之后，对试图进入大陆市场的台湾人和产品来说，曾经拥有的“不表态的灰色地带”，可能正逐渐被打破。