

数美数字风控行业系列白皮书

泛娱乐出海篇



前言

据 GSMA 数据显示，2021 年，全球移动互联网订阅用户已达到 42 亿。尽管全球宽带覆盖缺口已经从人口数的 1/3 减少至仅有 6%，但使用缺口人数仍有 32 亿，占了全球总人口 41%。

在 2022 年 3 月发布的第 49 次《中国互联网络发展状况统计报告》显示，截至 2021 年 12 月，我国网民规模达 10.32 亿，较 2020 年 12 月增长 4296 万，互联网普及率达 73.0%，增长率逐年放缓，从 2021 年 6 月起，手机网民占网民总数比例已达到 99.6%。

整体来看，国内移动互联网增长见顶，海外仍有大量新兴市场和成熟市场亟待开拓。本文在第一部分，分析全球人口和经济状况、互联网基建情况、用户付费意愿等，选取东南亚、中东和北美三大市场。在第二部分，对三大市场泛娱乐应用出海的机会，以及因政治、经济和文化差异导致的内容风险进行分析；第三部分将探讨出海企业风控体系搭建，包括视觉识别、文本识别和音频识别。

数美数字风控研究院致力于打造数字风控行业根据地，为从业者带来最前沿的产品、技术、政策解读及深度报告研究。汇聚专业力量，洞悉行业趋势。针对不同区域和行业，数美数字风控研究院将推出数字风控行业系列白皮书，泛娱乐出海篇为系列第一册。

版权说明

本册由数美数字风控研究院（公众号 ID: ishumei2015）编写和制作，版权归数美科技所有。

本册主要采用文献综述、桌面调研、行业访谈等调研方法，所涉及的图片、数据、参考文献、新闻报道均采集于网络公开信息，并已标注来源。

本册以分析海外市场潜力、展示海外市场内容风险、研讨海外内容审核能力搭建为主要目标，受公开资料和研究方法限制，本册内容和观点仅供读者参考和行业了解，不可视为企业提供的专业建议或指导。

本册为《数美数字风控行业系列白皮书——泛娱乐出海篇》第一版，如有修订，请关注“数字风控研究院”公众号获取新版。

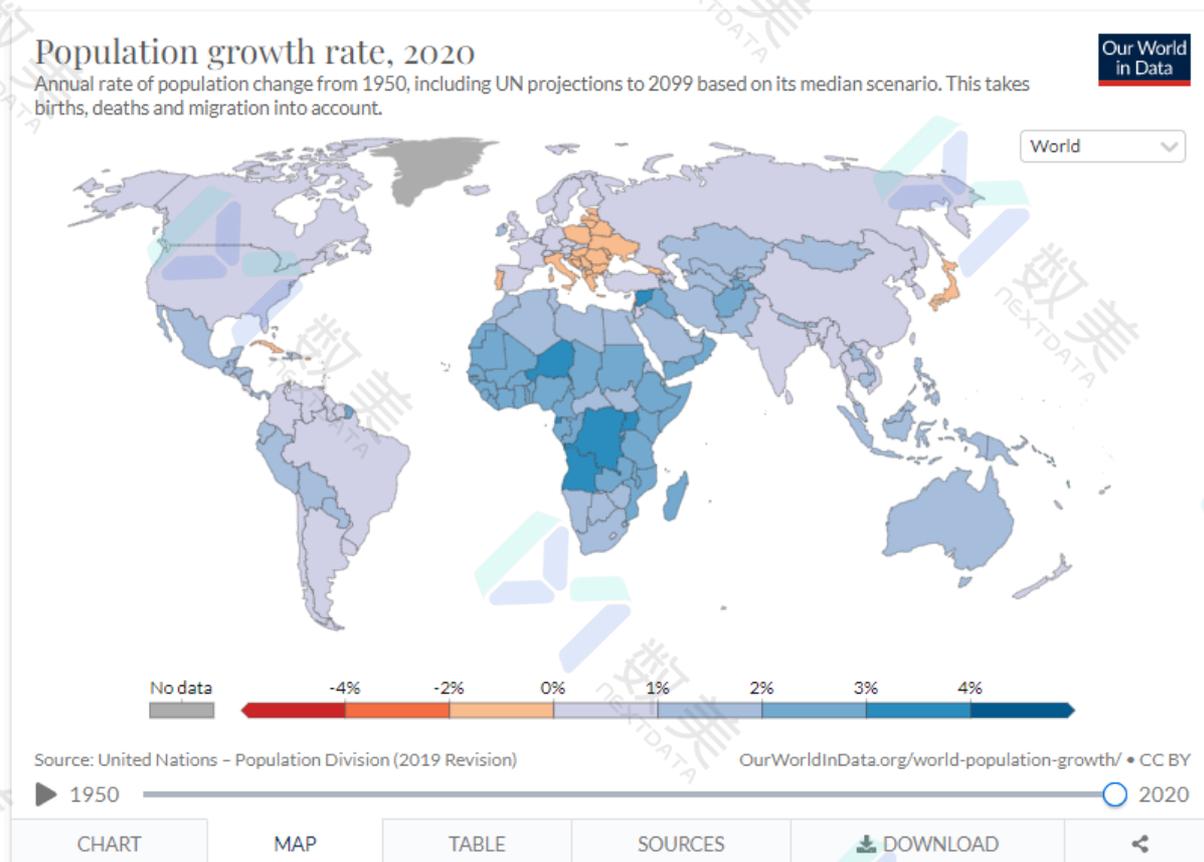
目录

前言	1
版权说明	2
一、全球互联网出海环境概览	4
1.1 全球市场分析	4
1.1.1 全球人口概况	4
1.1.2 全球互联网渗透率	7
1.1.3 付费意愿与付费能力	11
1.2 企业出海基本门槛	13
1.2.1 数据和隐私保护	13
1.2.2 海外应用商店准入	15
二、地区性风险分析	17
2.1 东南亚	17
2.1.1 地区市场特征	17
2.1.2 地区性风险	20
2.2 中东	23
2.2.1 地区市场特征	23
2.2.2 地区性风险	25
2.3 北美	30
2.3.1 地区市场特征	30
2.3.2 美国	32
2.3.3 加拿大	40
三、出海企业风控体系搭建	41
3.1 人工审核与智能审核	41
3.2 视觉识别	42
3.2.1 视觉识别思路阐述	42
3.2.2 视觉识别问题分类	43
3.3 文本识别	48
3.3.1 文本识别思路阐述	48
3.3.2 文本识别问题分类	48
3.4 音频识别	53
3.4.1 音频识别思路阐述	53
3.4.2 音频识别问题分类	54

一、全球互联网出海环境概览

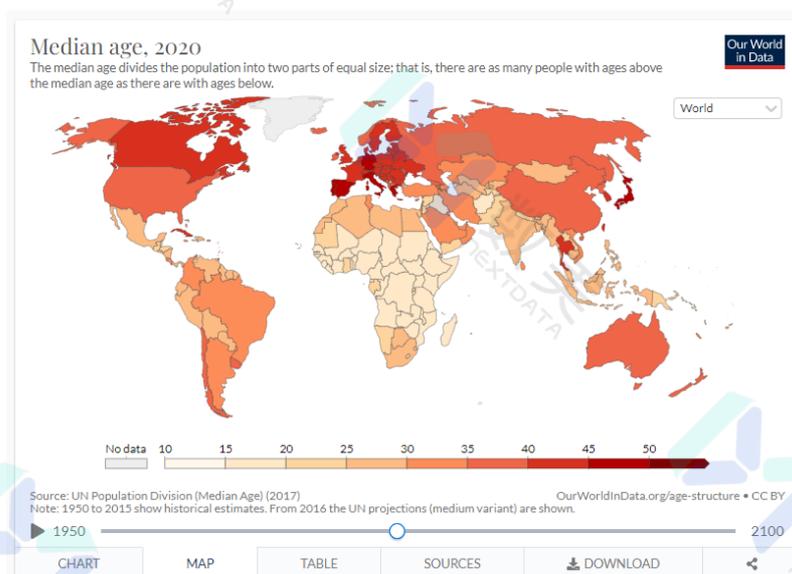
1.1 全球市场分析

1.1.1 全球人口概况



全球人口增长率概况
 数据来源：联合国

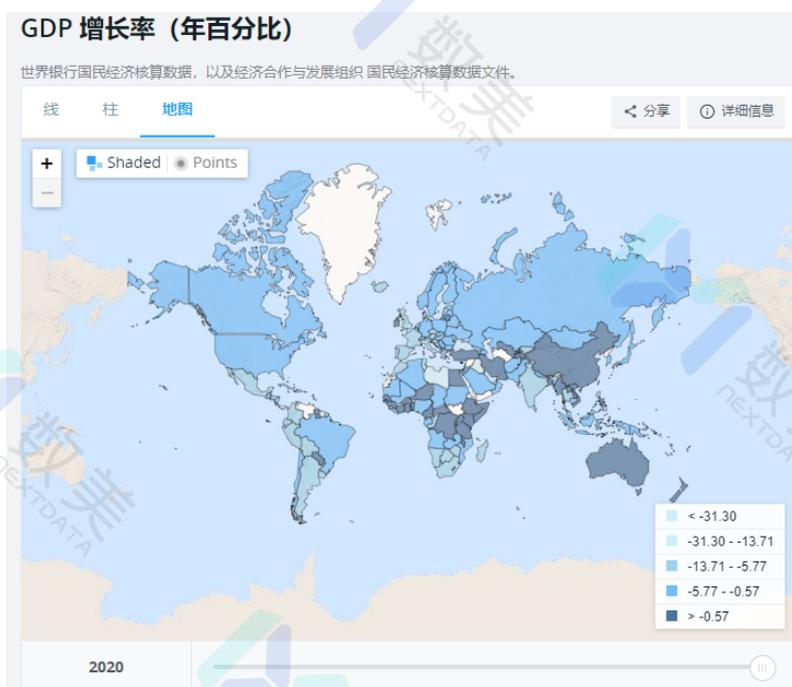
截至 2022 年 1 月，全球人口达到 79.1 亿，处于正增长状态的分别是北美、南美、东亚、俄罗斯等国家和地区，人口增长在 0-1%之间，中亚、中东、北非等国家则在 1-2%之间，撒哈拉以南的非洲增长高于 2%。人口增长迅速的国家，仍然有人口红利。



全球人口年龄中位数概况

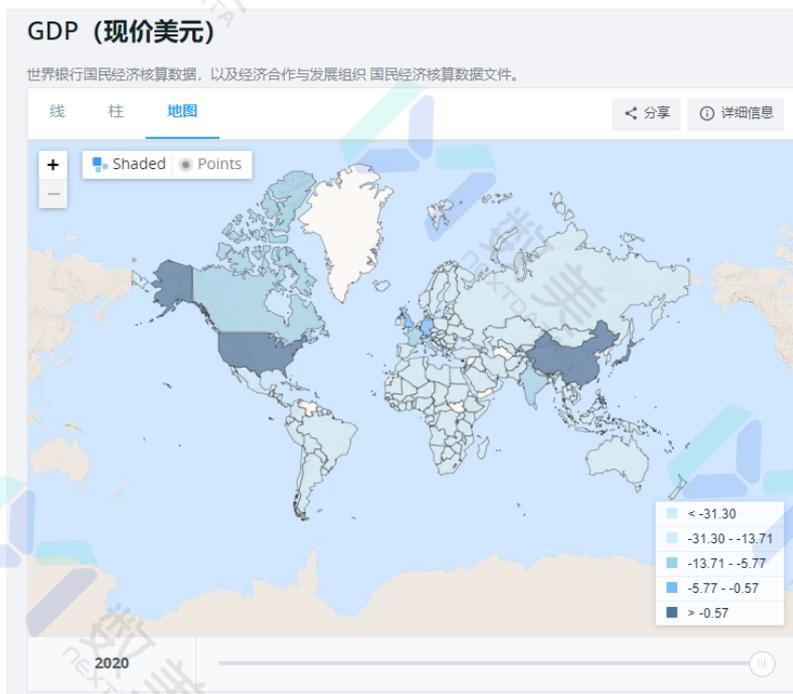
数据来源：联合国

全球年龄中位数为 30.9 岁，本报告将 25-35 岁的用户认定为有消费力且有增长潜力的群体，中位数处于这个区间的，集中在南亚、东南亚、中东、北非和南亚地区。



全球 GDP 增长率

数据来源：世界银行

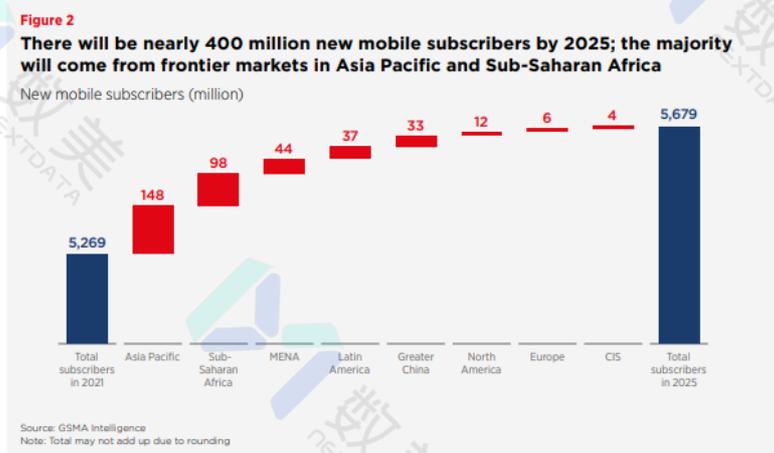


全球 GDP 总量分布
数据来源：世界银行

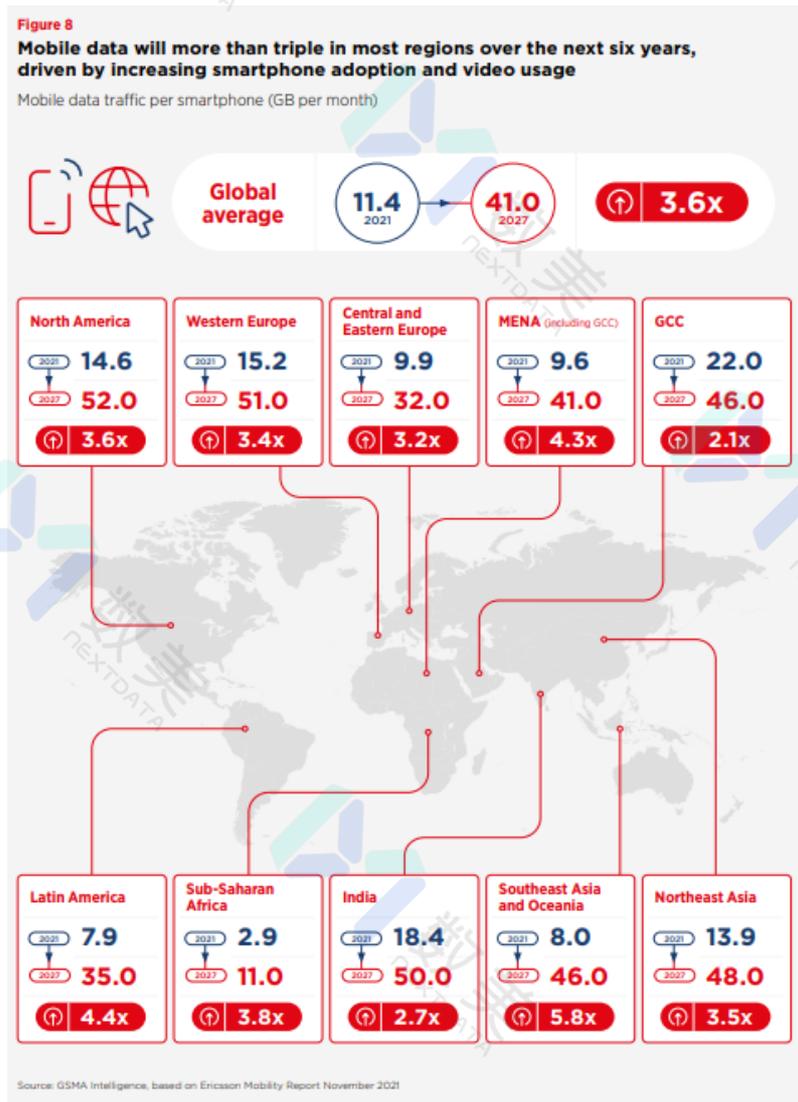
美国、中国、日本 GDP 总量排名前三，排名前十的除中国外依然集中在资本主义国家；从增长率（年百分比）来看，除中国外，中非、中亚和东南亚增速较快。

GDP 总量较高的国家，经济发展水平高，民众消费水平高，付费意愿强；GDP 增速较快的国家，增长潜力较大。

1.1.2 全球互联网渗透率



截至 2025 年移动互联网订阅新用户预测
数据来源：GSMA



未来 6 年全球移动数据增长分布
 数据来源：GSMA

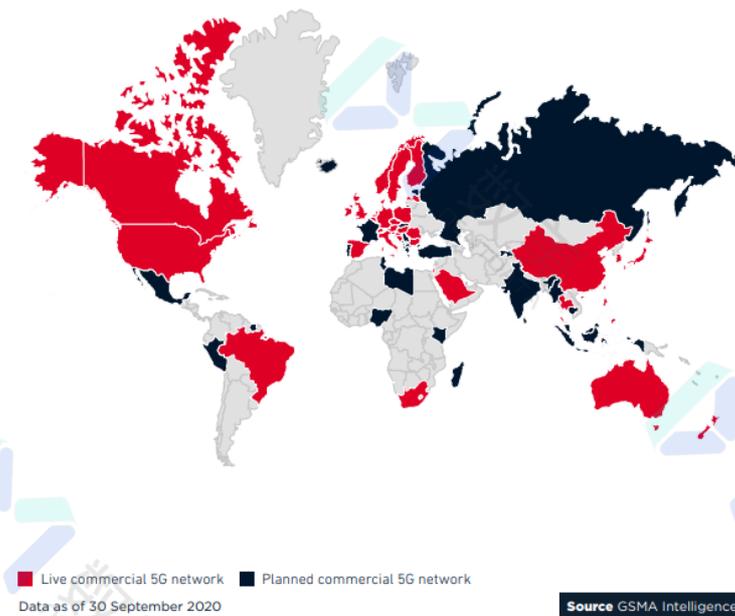
截至 2022 年 1 月，全球互联网用户达到 49.5 亿，互联网渗透率达到 62.5%¹。从移动端看，全球移动宽带覆盖人口已达到 94%。预计到 2025 年，全球移动用户将超 4 亿，新增部分主要来自于新兴市场，如亚太区和撒哈拉以南的非洲。

未来 6 年，全球移动数据量将翻两倍，东南亚增速最快，拉美和中东其次，北美和西欧地区，总量达到最高。

¹ We are social: digital 2022

[https://datareportal.com/global-digital-](https://datareportal.com/global-digital-overview#:~:text=4.95%20billion%20people%20around%20the,of%20the%20world's%20total%20population.)

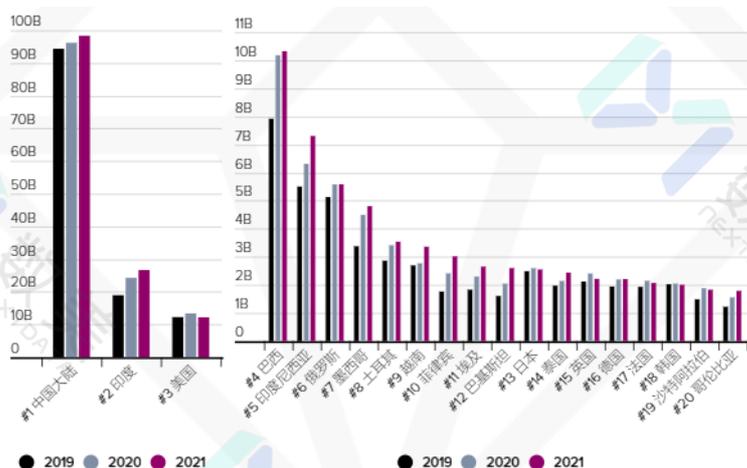
[overview#:~:text=4.95%20billion%20people%20around%20the,of%20the%20world's%20total%20population.](https://datareportal.com/global-digital-overview#:~:text=4.95%20billion%20people%20around%20the,of%20the%20world's%20total%20population.)



全球商用 5G 和计划商用 5G 铺设区域

来源: GSMA

尽管有疫情和相关经济约束,新的 5G 基建仍在大力推进中。已铺商用 5G 的国家,集中在北美、西欧、东亚、大洋洲和南美的巴西,而俄罗斯、东南亚、南亚、少数非洲国家和中美洲国家也计划铺设商用 5G。基建完备地区未来将持续具备更大的发展机会。



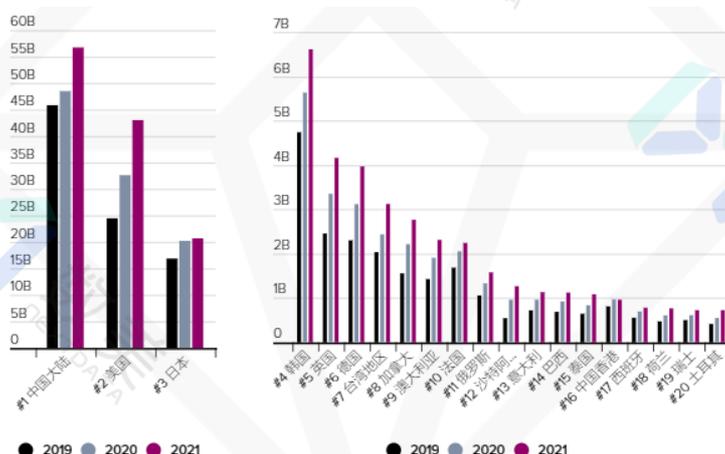
2021 年 TOP 20 热门移动市场下载量排名

数据来源: data.ai

在下载量增长方面,新兴市场占据主导地位,其中印度下载量表现突出。巴基斯坦、

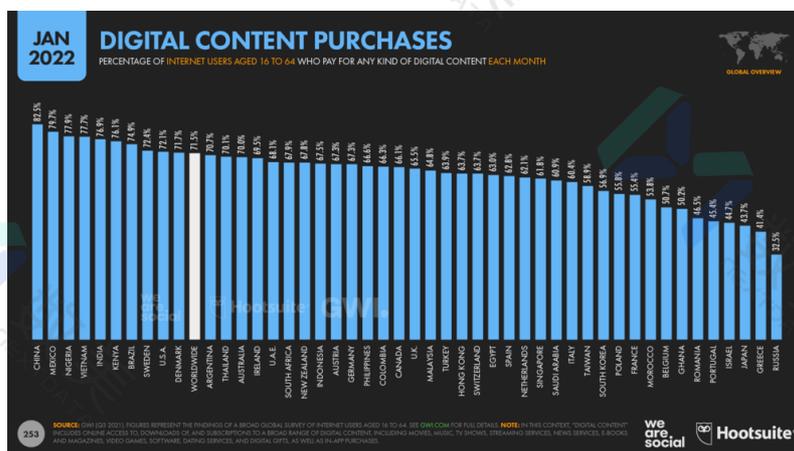
秘鲁、菲律宾、越南、印度尼西亚和埃及是下载量增长最快的市场，同比增长率分别为25%、25%、25%、20%、15%和15%。下载量增长快的区域，反映出当地有较大用户增长潜力。

1.1.3 付费意愿与付费能力



2021年TOP 20热门移动市场用户支出排名

数据来源: data.ai



全球用户每月数字内容支出排名

数据来源: We are social

用户支出代表了网民的消费水平和付费习惯，直接影响应用的未来收益潜力，中国因人口基数庞大、经济发展增速快，用户支出占据首位。

从 Top 20 热门移动市场的用户支出表现看，发达国家如美国、日本、韩国、英国，

用户支出排名靠前。其中，因移动游戏和应用内订阅成为主流，美国的用户支出实现大幅增长，达到 430 亿美元，与 2020 年相比增加 104 亿美元，相当于同比增长率为 30%。

综合人口红利、经济发展速度和经济水平、市场成熟度三方面，本文选取东南亚、中东和北美作为典型区域重点讨论。

1.2 企业出海基本门槛

1.2.1 数据和隐私保护

2020 年爆发的新冠疫情给全球经济带来重创。在全球经济增长减缓背景下，数字经济反而逆势增长。据信通院 2021 年 8 月发布的《全球数字经济白皮书——疫情冲击下的复苏新曙光》²显示，2020 年，测算的 47 个国家数字经济增加值规模达到 32.6 万亿美元，占 GDP 比重为 43.7%，数字经济成为全球经济增长新动能。

当数字平台成为访问商品和服务的唯一工具，如何增强消费者对数字经济和电子商务的信任，成为亟需解决的问题。越来越多的国家开始将数据保护与隐私纳入立法进程。2018 年 5 月 25 日，《通用数据保护条例》（GDPR）在经过两年的缓冲期后正式步入执法阶段。作为全球首部全面的个人数据保护法，它的出台极大地影响了欧洲乃至全世界的数据保护立法进程。

就在 GDPR 正式实施不久，美国加州便匆忙通过了 CCPA 草案，而我国于 2021 年实施的《个人信息保护法》与《数据保护法》在制定过程中也参考了 GDPR 的个人数据保护框架。

据联合国贸发组织的最新数据显示，截至 2021 年 12 月 14 日，在 194 个联合国会员国中，有 71% 的国家/地区已有数据保护与隐私立法，9% 已经制定了相关法律草案。由于各国在立法和监管进度、方式存在差异，数据和隐私保护已成为企业出海合规不可忽视的法律风险。

² 2021.06，来源：信通院，《全球数字经济白皮书——疫情冲击下的复苏新曙光》，

<http://www.caict.ac.cn/kxyj/qwfb/bps/202108/P020210913403798893557.pdf>

全球隐私法规一览

国家	隐私法规
欧洲	《通用数据保护条例》(GDPR) 《非个人数据自由流动条例》
美国	联邦立法:《公平信用报告法》、《健康保险流通和责任法》、《金融服务现代化法案》、《儿童在线隐私保护法》、《金融消费者保护法》 州立法:《加州消费者隐私法案》(CCPA)、《纽约州禁止黑客和改善电子数据安全法案》、《弗吉尼亚州消费者数据保护法》、《科罗拉多州隐私法案》
中国	《网络安全法》、《数据安全法》、《个人信息保护法》
韩国	《个人信息保护法》
日本	《个人信息保护法》
泰国	《个人数据保护法》
马来西亚	《2010年个人数据保护法令》
新加坡	《个人信息保护法案》
印度尼西亚	2008年关于电子信息和交易的第11号法 2019年关于电子系统和交易运行的第71号政府条例 2016年关于电子系统中个人数据保护的通讯及资讯管理部第20号条例(部颁条例第20号)
菲律宾	《数据隐私案》
老挝	《电子数据保护法》
越南	《网络安全信息法》

1.2.2 海外应用商店准入

除了各国的数据和隐私合规,应用商店的应用上架规则以及审核指南成为 APP 出海的又一准入条件。海外市场,App Store 和 Google Play 在 iOS 和安卓设备上的安装比例分别超过 95%,是国外 APP 上架的主流应用商店。据广告分析公司 Pixalate 报告显示,2021 年上半年,App Store 和 Google Play 共下架应用超过 81.3 万款,占比超过 15%³,应用商店也因此承担着“守门人”的角色。

³ 2021.09.14, 来源: Pixalate, Pixalate releases H1 2021 delisted mobile apps report,

<https://www.pixalate.com/blog/delisted-apps-report-h1-2021>

Google Play⁴

Google Play 应用商店主要通过开发者分发协议（Developer Distribution Agreement）和开发计划政策（Developer Program Policy）来规范上架 APP。Google Play 的政策是随趋势更新的，每年至少有两次较大的更新⁵。

开发者分发协议是 APP 开发者与 Google Play 签订的一份具有法律约束力的合约，它规范了开发者的发布行为和推广商业行为。通过签订协议，APP 开发者同意委托谷歌在 Google Play 中分发产品，并承担相应的责任。分发协议中明确提到，谷歌虽然不承担监控产品或其内容的义务，但一旦发现某款产品违反法律或违反谷歌政策，谷歌有权下架该产品。

开发者计划政策则是对 APP 内容的具体规定，包括受限内容，假冒行为，知识产权，隐私权、欺骗行为和设备滥用，获利和广告，商品详情和宣传，垃圾内容和最基本的功能，恶意软件，移动垃圾软件，家庭计划等。

Google Play 按照违规严重程度设置惩罚措施，且会对违规行为进行累计。例如，APP 中的 UGC 内容中含有暴露图片被视为轻度违规，在第一次和第二次发现后，Google Play 会向开发者发出警告，累计多次后也会导致产品下架或账号被封。广告欺诈、干扰性广告、用户安全问题以及儿童安全都被视作严重的违规行为，发现一次立即下架。此外，如果一个账户被封，所有关联账户都可能被封杀。

App Store⁶

与 Google Play 一样，苹果 App Store 也为开发者设置了一套针对技术、内容和设计方面的标准——App Store 审核指南（App Store Review Guidelines）。App Store 的审核团队目前拥有超过 500 名审核员，可覆盖 81 种语言的 APP 审核，每位审核员每周审核 200 款 APP，单个 APP 的审核时长为 12 分钟。2017 年至 2019 年，App Store

⁴ 谷歌开发者政策中心网址：<https://play.google.com/intl/zh-CN/about/developer-content-policy/>

⁵ 2020.04.17，来源：Enjoy 出海，Google Play 政策运营专家亲谈：游戏出海，开发者要注意哪些常见违规问题？
<https://enjoyglobal.net/detail/news/LNews371>

⁶ App Store 审核指南中文版官方网址：<https://developer.apple.com/cn/app-store/review/guidelines/#objectionable-content>

收到的 APP 申请中有 33%-36% 会被拒绝⁷，将近 40% 的拒审率也体现了 App Store 严格的审核标准。在向 App Store 提交应用前，开发者应对照审核指南自查自纠，避免审核不通过造成应用无法准时上线。

App Store 审核指南从安全、性能、商务、设计、法律等五方面详细介绍了 APP 审核细则。内容安全一直是 App Store 审核的重要内容之一，在第一章“安全”中首先介绍了 App Store 如何审核“令人反感的内容”和“用户生成的内容（UGC）”。

审核指南规定，APP 中不应包含令人反感的内容，即“不应包含具有攻击性、不顾及他人感受、令人不安、惹人厌恶、低俗不堪或只是让人感到毛骨悚谈的内容”。例如：

针对宗教、种族、性取向等进行诽谤、歧视；对人或动物遭到杀害、残害、酷刑、虐待的写实描述；鼓励暴力；鼓励非法使用或不负责地使用武器和危险物品的描述，或者促进军火或弹药购买的描述；公然宣传黄色或色情内容的材料，包括其中含有色轻内容或是为卖淫活动提供便利的一夜情约会 APP；具有煽动性的宗教评论，或者对宗教文本进行错误或误导性的引用；虚假信息和功能，其中包括不准确的设备数据或用于恶作剧/开玩笑的功能，如虚假的位置跟踪器。

由于可能会产生知识产权侵权、匿名欺凌等难题，审核指南还规定，有 UGC 功能的 APP 应设置内容审核机制和举报机制，从而避免 APP 被滥用。

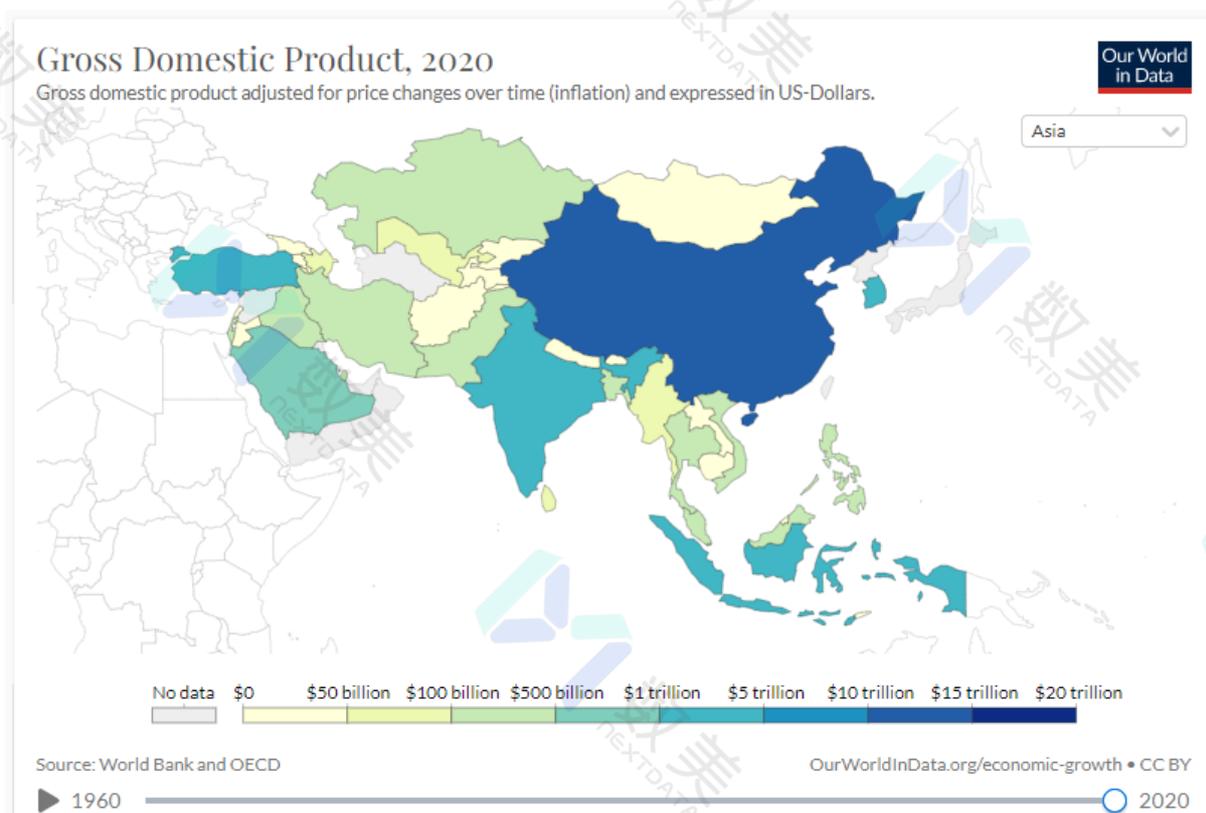
Google Play 和 App Store 作为绝大多数海外 APP 的上架入口，在 APP 的内容审核上握有仅次于政府立法的话语权。相比政府立法，应用商店自己制定的协议更具灵活性，它可以及时更新，不必受繁冗的立法程序所累；同时，由于应用商店本身就是技术力量雄厚的互联网公司，它们能够及时以技术方案识别、处理上架 APP 的内容问题，比政府执法更为直接有效。

⁷ 2021.06.06，来源：稀土掘金，解密苹果应用审核团队，<https://juejin.cn/post/6970363897668698148>

二、地区性风险分析

2.1 东南亚

2.1.1 地区市场特征



亚洲地区 GDP 总量分布
数据来源：世界银行

东南亚区域一共有 11 个国家，总人口超 6.7 亿，全球占比 8.58%，是亚洲排名第三的区域。其中，人口排名靠前的分别是印度尼西亚、菲律宾、越南、泰国、缅甸。按人均 GDP 水平考虑，排名靠前的分别是新加坡、印尼、泰国、菲律宾、马来西亚。（数据来源：联合国）

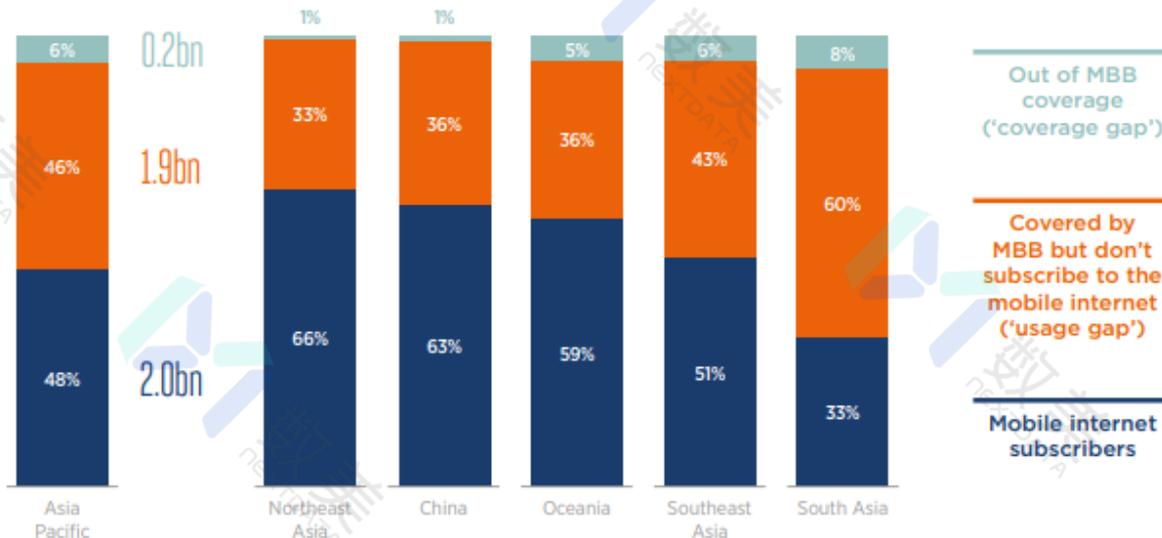
考虑到影响出海的人口和经济水平两项指标，本文主要讨论印度尼西亚、菲律宾、越南、泰国、缅甸、新加坡六国。

Figure 21

Source: GSMA Intelligence

Around 2.2 billion people across Asia Pacific don't use mobile internet; for most, coverage is not the issue

% of population (2019)



Note: totals may not add up due to rounding

亚洲地区移动互联网覆盖情况人群占比

图表来源: GSMA

入网情况:

谷歌、淡马锡和贝恩公司联合发布的《2021年东南亚数字经济报告》显示。2021年，新加坡、马来西亚、印尼、菲律宾、泰国、越南六国新增4000万互联网用户，同比增长10%，总数达4.4亿，六国已有75%的人口接入互联网。但移动入网占比在整个亚太地区排名倒数第二，仅有51%用户入网，仍有较大增长空间。

智能手机出货量:

根据GSMA数据，预计到2025年，全球智能机市场前三分别是中国、印度和印度尼西亚，其中印尼出货量将达到3.5亿，东南亚地区排名第一。

消费意愿:

据data.ai测算，全球二十大用户支出热门市场里，东南亚国家未被列入其中。受经济发展水平影响，东南亚地区还处于互联网发展早期阶段，以积累用户为主。

整体而言，东南亚地区尚处于互联网发展早期阶段，属于新兴市场。用户增长空间大，地缘接近，同样受到东亚流行文化影响，有天然的出海落地优势。

2.1.2 地区性风险

1) 特殊禁忌

这里主要指皇室、政治等相关话题，在泰国尤其需要注意。

泰国作为君主制国家，皇室地位特殊，涉及国王或皇室的内容需要尤其注意。普通人不能使用一般的词汇形容王室，而应该用特殊的“王语”，例如，“教训”一词在王语中称为“圣训”或者“王训”，黄色是皇室的颜色，需避免污蔑性的内容。

除此外，泰国还有专门立法禁止亵渎王室与君主制。其中，《计算机犯罪法案》禁止上传虚假的或影响国家安全的信息，也被用于起诉涉嫌诋毁王室的违法行为；刑法规定，诽谤、侮辱或威胁国王、王后、继承人或摄政王的人将被判处 3 至 15 年有期徒刑。

越南在 2019 年出台了《网络安全法》，认定在网络批评政府属于犯法行为，网络提供商需要按要求提交用户数据。

2) 宗教因素

东南亚宗教环境比较复杂，伊斯兰教和佛教占比最高。

伊斯兰教是最大的宗教，40%的当地人口为穆斯林，约 2.4 亿人，主要分布在印度尼西亚、文莱、马来西亚、泰国南部和菲律宾南部；38%的当地人口为佛教徒，约 2.05 亿，主要分布在越南、泰国、老挝、柬埔寨、缅甸和新加坡。基督教为菲律宾、印度尼西亚东部、马来西亚东部及东帝汶的强势宗教，其中，菲律宾拥有全亚洲最大的罗马天主教人口。

印尼对互联网上的色情与反伊斯兰内容尤其敏感。政府会通过《电子信息法》等法律和监管框架以及与 ISP（互联网服务提供商）和网吧的合作来对互联网内容进行管理 & 监控。

马来西亚也是穆斯林比较集中的国家，根据马来西亚《通讯及多媒体内容准则》等相关法规规定，互联网及游戏产品中不允许出现算命、猪肉、赌博等成分，对暴力、色情等成分也同样敏感，

东南亚六国主要宗教

国家	主要宗教
印度尼西亚	伊斯兰教、基督新教、罗马天主教
菲律宾	罗马天主教、伊斯兰教、基督新教
越南	越南民间信仰及无宗教、佛教、罗马天主教
泰国	上座部佛教、伊斯兰教、基督教
马来西亚	伊斯兰教、佛教、基督教
新加坡	佛教、伊斯兰教、基督教

3) 未成年人

东南亚六国中，新加坡的互联网未成年人保护相对较为完善。

新加坡的刑法、《电影法》和《不良出版物法令》等规定，以电子方式传播淫秽影片、书籍等不良内容的个人将被处以监禁和罚款；如果是向未成年人传播淫秽内容或者诱使、教唆未成年人实施上述行为的，刑罚加重。新加坡的互联网服务提供商必须确保所提供的内容符合《互联网行为准则》，限制一些对儿童有害内容的网站。父母还可以订阅互联网过滤服务，保护孩子安全上网。

4) 色情/性别话题

各国对于色情内容的监管都有相对成熟和严格的立法。

新加坡的《色情电影法（Pornographic Films）》规定，创作色情影片用于展示和传播，最高处以 4 万美元罚款或 2 年监禁；传播或持有用于传播的色情电影，最高处以 8 万美元罚款和 2 年监禁。

印尼在 2008 年通过《反色情法》，这一法案一直以来受到较大争议，但获得很多伊斯兰教徒的支持。“禁止公民生产、制作、复制、分发、广播、进口、出口、提供、买卖、租借色情内容”。

菲律宾则有严重的儿童性犯罪问题，联合国在 2016 年曾发布《远离伤害：应对菲律宾儿童网络性虐待问题》，呼吁保护儿童，起诉施虐者。《9775 号共和国法》对儿童色情犯罪有明确规定，制作、传播和商业持有儿童色情影片都属于违法行为。同时规定，

所有互联网提供商在发现任何形式的儿童色情内容时，需要在在 7 天内通知相关机构。

泰国刑法第 287 条规定，用于交易的成人内容的生产、分发和持有都属于违法行为，最高处以 3 年监禁，持有儿童（18 岁以下）色情影片的人将被处以最高 5 年监禁或 10 万泰铢罚款。

马来西亚政府在 1998 年发布的《通信与多媒体法令》上，起草了一项《内容准则》，为所有的互联网内容提供者列出了整套的指导方针、程序和好的实践。规定进一步对被禁止的互联网材料及相关行为做出了分类，要求网络平台不得含有不雅内容、淫秽内容、暴力、恐吓内容、脏话、虚假内容、儿童（不宜的？）相关内容、家庭价值观以及与残障人士相关内容。

5) 恐怖暴力

东南亚地区，由于复杂的宗教和政治局势，极端言论和事件频发，平台尤其需要注意。

2017 年，因 Telegram 上充斥着大量极端主义思想，宣传仇恨、恐怖主义与指导炸弹制作方法，印度尼西亚政府正式发布封锁社交网络 Telegram 应用程序的通告。随后一年，印尼通过《反恐法案》，除了相较此前一版，增加预防、侦查、监控、恐怖分子改造等方面内容，网络空间的暴恐言论监管也得到进一步强化。

6) 仇恨言论

近 10 年，仇恨言论在全球都有愈演愈烈的趋势。

印度尼西亚刑法第 156 条规定，对任何别的民族的仇恨言论，最高可处以 4 年监禁和 300 卢比罚款。

新加坡在 2018 年颁布《广播法》和《互联网操作规则》。《广播法》规定了互联网管理的主体范围和分类许可制度，《互联网操作规则》则明确规定了互联网服务提供者和内容提供商应承担自审内容或配合政府要求的责任。1996 年发布的《互联网操作规则》第四条第一款规定，禁止播发“有违公共利益、公共道德、公共秩序、公共安全、国家和谐以及新加坡法律所禁止的其他内容”。这两部法规成为新加坡互联网管理的基础性法规。

2.2 中东

2.2.1 地区市场特征

提及中东，一部分人会想到石油和遍地富豪，一部分人会想到神秘的伊斯兰文化。经济与技术的“现代”与文化的“传统”在这片地区深度交融，在移动互联网建设的逐步完善下，中东地区逐步激发出巨大的互联网消费需求，电商、社交、音视频等新兴互联网行业成为中东市场的新热点，而中东也成为了出海的新蓝海市场。

卡塔尔、阿联酋、科威特、沙特阿拉伯、巴林和阿曼这六个国家是海湾阿拉伯国家合作委员会——海合会（GCC）的成员，这六个国家也是中东地区消费能力最强的海湾阿拉伯国家。这里的主要人口为信奉伊斯兰教的阿拉伯人，阿拉伯语是占绝对地位的主要语言。



海合会成员国

图片来源：海合会官方网站 <https://www.gcc-sg.org>

而海湾地区的六国也体现出明显的高收入趋势，根据世界银行数据显示，在 2020 年各国的 GDP 情况：

2020 年海合会六国 GDP 数据表

收入情况	卡塔尔	阿联酋	科威特	沙特阿拉伯	巴林	阿曼
总 GDP (亿美元)	1444.1	3588.7	1059.6	7001.2	347.3	739.7
人均 GDP (万美元)	5.0	3.6	2.5	2.0	2.0	1.4

网络普及率：

阿联酋在 5G 部署等网络方面处于世界领先水平，网络普及率高达 98%，是中东地区互联网普及率最高的国家。

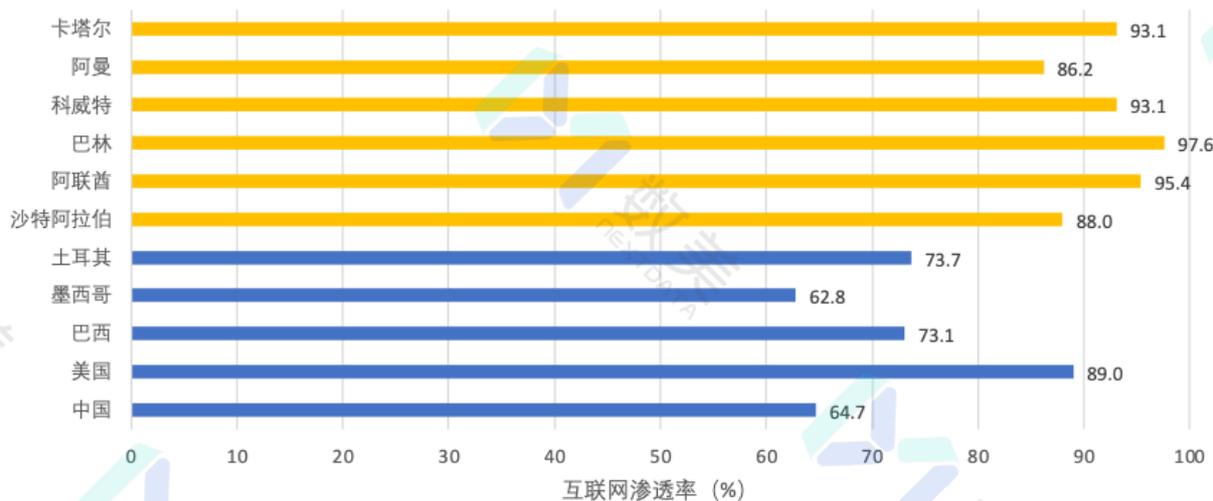
2020 年，科威特的网络渗透率达到 99%，全国拥有 420 多万名网络用户。

2020 年，作为中东第一大经济体和最大消费市场的沙特阿拉伯网络用户数量达到 3223 万，网络渗透率达到 93%。

整体上，中东地区约 5 亿的总人口中互联网用户可达近 3 亿，总体互联网渗透率可达 58.7%，这一数据高于世界平均水平。

网民年龄：

按年龄分析，25 岁以下人口在中东地区占比约 50%，具有巨大的年轻人口红利。



世界部分国家互联网渗透率

数据来源：Statista, AppAnnie, We are social & Hootsuite

消费意愿

中东地区体现出极高的消费热情和消费能力。根据折扣网站 Picodi 在 2020 年的一份报告指出，每名沙特阿拉伯女性平均每年会在化妆品上花费 909 美元，这一数值接近英国女性消费（505 美元）的两倍，是中国女性平均水品（约 100 美元）的九倍多。根据中东地区极具竞争力的出海电商品牌 Shein 统计，中东用户在 Shein 平台的单笔消费金额经常能够达到 200 美元以上，平均客单价高于在美水平的 150 美元。

以上种种因素将中东塑造成为一个年轻、富有、充满活力的市场，与“圈地跑马”时代的中国市场拥有极高的相似度，在移动互联网市场的各个细分行业中都有巨大的开发潜力和潜在市场规模。

2.2.2 地区性风险

1) 特殊禁忌（涉政）

在 MENA（中东与北非）地区，仍然有多种政治体制存在，例如实行共和制的埃及、黎巴嫩、巴勒斯坦、也门、阿尔及利亚等，实行君主制的约旦、沙特阿拉伯、阿曼、科威特、巴林、卡塔尔、摩洛哥。

政治体制的不同会产生新的内容风险，在实行君主制的国家存在王室/皇室，是国家的直接统治者。因此在公开场合或网络社交平台谈论或议论皇室在很多国家会是约定俗成的禁忌，至于抹黑皇室形象、涂抹和玷污皇室人员肖像的行为更有可能面临法律的制裁，社交平台在进入中东地区的君主制国家时，就需要对该国皇室成员的名字和肖像有所了解，提高政治敏感度。更有甚者，还要对皇室更替、成员斗争有舆情上的关注，以免自身平台沦为舆论斗争工具。

2) 宗教因素

中东地区的大多数国家或是以伊斯兰教为国教，或是有很高比例的人口信仰伊斯兰教，整个地区都广泛受到伊斯兰文化的影响。而这种影响非常深远，例如游戏中也有专门避免敏感因素的设计：女性角色必须要让衣服覆盖全身，不能露出胳膊等“羞体”；动物设计禁止出现猪，食物系统禁止出现酒；美术设计和装饰设计不能出现十字架等敏感元素。

在伊斯兰教的宗教相关方面，古兰经以及宗教刊物；清真寺、伊斯兰寺庙以及猪形象、猪卡通图案、猪肉、猪皮等涉及猪元素的都属于敏感元素，都是要格外注意的。对于社交平台，首先对广场内容要重点清查，谨防相关动物形象的出现，如果有向电商方面拓展的规划，那么也要注意上架商品、商品图、商品名的审核。



中东地区禁忌的六芒星、卐和十字架等部分符号
图片来源于网络

另外，在中东地区其他宗教的符号和象征也是严格管控的形象，例如犹太教的六角星、纳粹党“卐”字标、基督教的十字架以及佛教、印度教、道教、撒旦教、共济会等其他宗教的相关图案，以及耶稣、圣诞老人、佛像等具有宗教象征意义的人物形象，带有相关符号的商品在中东地区也严格禁止上架。值得一提的是，动漫迷熟悉的 DC 漫改游戏《不义联盟》（英文名《The God Among Us》）就因含有“God”单词在阿联酋被封禁，即便是改名《The Mighty Among Us》也未能幸免于难，仍以游戏内包含“God”单词被封。

3) 色情因素

中东地区受伊斯兰文化影响，在色情问题上十分敏感，同时这种敏感甚至跨越了“色情”的维度，成为了社会生活和文化的一种约束。例如国人比较熟悉的部分中东地区国家女性需要着黑袍出门，只能裸露双眼在外的规定，就能够折射出中东地区在性问题和开放性上的保守程度。



中东地区着黑袍出门的女性
图片来源于网络

在伊斯兰文化中有一个概念叫做“羞体”，即人必须覆盖不可示人的躯体。随着经济和文化的发展，“羞体”的定义其实已经呈现出了历史性和地区性的差异。例如在国内，男性“羞体”一般指肚脐到膝盖部分的躯体，女性“羞体”一般指除去面部、手部小臂等肢体外的躯体（包含头发）。所以能够看到国内穆斯林女性往往着头巾遮盖头发，但面部没有遮挡。



穿着现代的黎巴嫩女性
图片来源于网络

这种差异在中东地区的穆斯林群体中其实也存在，有的国家要求女性着黑袍且只能裸露双眼，有的国家则要求遮盖头发和身体但可以裸露面部，这其实是宗教、经济、思想文化互相博弈话语权产生出的多种多样的社会形态。甚至在一些经济水平很高、思想解放程度高但宗教影响力强的地区，出现了女性虽着黑袍出门，但黑袍下却有牛仔裤和高跟鞋的奇特现状。也有电商数据显示，泳衣、内衣乃至情趣内衣在中东市场的销量表现都比较亮眼，但因为要避开女性三角区和胸臀背等区域，商品详情页不能用人体模型或模特，直接展示商品反而更安全。至于像黎巴嫩这样区位上靠近欧洲的国家则更加开放。

中东地区对色情问题的敏感度实则是伊斯兰文化影响力的差异，但总体上仍是趋于严格的。低俗、性暗示、同性恋相关的内容，甚至涉及“情人节”元素的内容都是十分敏感的。这种严格不仅仅约束色情内容，对男女着装、男女裸露问题都有辐射。2015年，索尼中东就曾宣布，PS平台游戏《教团：1886》因男女裸露、含有妓院建筑设计等问题被沙特阿拉伯和科威特封禁。

4) 恐怖暴力

提及恐怖暴力问题，可以说是整个中东地区的一块“逆鳞”。历史上，曾经在中东地区诞生出三大臭名昭著的恐怖组织：基地组织、塔利班、伊斯兰国（ISIS）。

而随着时代的发展，恐怖组织和恐怖主义也加大了对媒介的关注，从游击队式的招募到占领地方电视台洗脑，再到全面进驻社交媒体进行全球散播，防范恐怖主义已经真实成为了中东地区社交媒体的重要责任。



因账号被处理，ISIS 发布对 Twitter 联合创始人 Jack Dorsey 的威胁信息
图片来源于网络

据报道，恐怖组织伊斯兰国 ISIS 就十分善用社交媒体作为宣传工具，2014 年-2015 年间，Twitter 删除了 12.5 万个威胁发动或宣传恐怖行动的账号，其中的 2.1 万个是 ISIS 组织仅用一月便新建完成的。ISIS 组织通过全球社交媒体迷惑大量支持者，甚至曾成功通过 Facebook 招募到澳洲的一名 17 岁少年。

目前，虽然恐怖组织 ISIS 已经于 2019 年被宣告消灭，但中东地区并不认为恐怖主义和恐怖分子被完全消灭。中东各国也对恐怖和暴力相关内容、产品有非常严苛的管控红线。例如模拟枪、子弹等造型的产品，包含军绿、土黄迷彩元素的服装，遥控飞机等玩具乃至望远镜、无线对讲机等都在中东的消费市场有不同程度的管控政策。相应地，这些物品出现在中东社交平台时也具有极高的风险。



ISIS 组织黑旗
图片来源于网络

此外，还要谨防宣扬恐怖主义的相关内容，例如恐怖组织旗帜和符号、包含枪械和血腥场面的图片以及宣扬“圣战”和美化恐怖袭击的相关言论，以上内容可能出现在用户头像、简介、广场内容等场景。此外，带有直播和通话功能的实时社交平台也要谨防恐怖主义宣扬内容以画面背景和音频的方式混入其中。

2.3 北美

2.3.1 地区市场特征

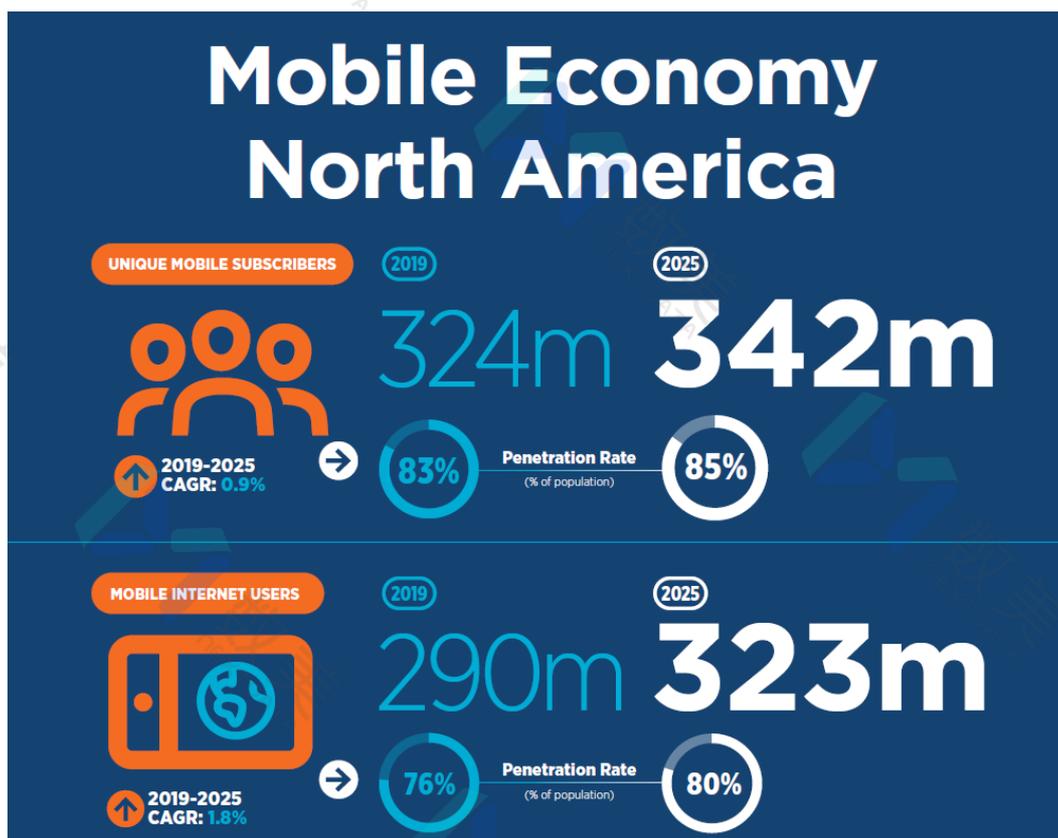
北美地区一般由两个国家和三个地区组成：美国、加拿大、百慕大、格陵兰以及圣皮埃尔和密克隆。据世界银行数据，美国 2020 年 GDP 为 20.953 万亿美元，加拿大 1.645 万亿美元，而另外三个地区的人口均不足十万。从市场规模和人口规模来看，美国和加拿大是北美地区的绝对主力。

互联网基础设施完备

2019 年，北美地区的 4G 普及率超过 80%。近几年，北美也开始加快部署 5G，预计 2025 年 5G 的普及率将首次超过 4G。

网络普及率高

据 GSMA 2020 北美移动经济报告数据，到 2025 年，北美移动用户将超过 3.4 亿人，占总人口的 85%；移动网络用户数量将超过 3.2 亿人，普及率达 80%。



北美移动用户和移动网络用户数量
 图表来源：GSMA 2020 北美移动经济报告

线上娱乐习惯已养成

由于拥有完备的网络基础措施，高网速低资费也使得北美用户使用线上娱乐服务更方便。2019年，41%的北美用户付费线上观看电视节目，将近60%的美国和加拿大用户会在手机上玩免费游戏。

2.3.2 美国

1) 美国网络内容监管中坚力量：行业自律

与欧盟等地区通过政府立法监管网络内容不同，美国的网络内容主要依托行业合作进行自我管理。

美国政府在网络内容监管上的步伐较为谨慎，其著名的第230条法规规定，互联网服务的运营商无需对平台上的第三方言论承担法律责任，这几乎等同于是网络平台的豁免令。虽然美国近年来就废除230条款一事展开多轮辩论，但截至目前，互联网平台无需为平台内容负责的立法现状仍未改变。



特朗普下令重审第 230 条

图片来源：FT 中文网

然而，政府立法监管的缺位并不代表美国网络世界毫无限制。其一，美国政府虽然囿于对自由言论的保护，无法通过立法约束平台，但美国国会也在不断向企业施压，要求企业采取措施解决网络内容乱象。其二，在美国活跃的社交平台基本上都属于全球头部企业，虽然美国本土无立法限制，但其他地方的立法要求，如欧盟，也会倒逼美国企业对网络内容进行限制。其三，如前所述，海外的主流应用商店 Google Play 和 App Store 均为上架 APP 设置了内容审核条款，头部行业自发设置的准入门槛成为美国互联网企业不得不遵守的硬性要求。

接下来，本篇将针对美国网络内容监管的重点领域进行分析。

2) 未成年人

20 世纪 90 年代，由于未成年人容易接触到互联网上大量的色情内容，引发民众与政府的担忧与抵制，由此开始了美国第一波互联网管制浪潮。此后，美国联邦多次尝试通过立法建立一套强制内容审核系统，但由于《美国宪法第一修正案》对言论自由的强力保护，几次提案都以失败告终。几次比较著名的立法提案包括 1996 年《通讯规范法案》(Communications Decency Act, CDA)、1998 年《儿童在线保护法案》(Child Online Protection Act, COPA)、2000 年《儿童互联网保护法案》(Child Internet Protection Act) 以及 2000 年《儿童在线隐私保护法案》(Child Online Privacy Protection Act, COPPA)。

《通讯规范法案》(CDA)

《通讯规范法案》是美国首次尝试通过立法管制未成年人接触互联网色情内容。1996年,美国通过了被称为是“美国新旧通信世界的分水岭”的《电信法》(Telecommunications Act of 1996, TA),《通信规范法》是《电信法》第五章的内容,主要目的在于禁止通过网络向未满十八岁的未成年人传输“低俗(Indecent)、淫秽(Obscene)”内容。

《通讯规范法案》包含两部分,第一部分规定,通过互联网向不满十八岁的未成年人提供或者传送下流、淫秽的评论、要求、建议、形象等材料或信息,情节严重的应当承担刑事责任。其第二部分则为著名的第230条,规定互联网服务的运营商无需对平台上的第三方言论承担法律责任。但在1997年的雷诺诉美国公民自由联盟案中,美国最高法院裁定,法案的第一部分违宪,限制未成年人接触网络低俗淫秽的法条被判无效。

最高法院判定《通讯规范法案》反低俗内容条款违宪的原因主要是法案中对“低俗”、“淫秽”的定义表述模糊不清。作为一部旨在对言论内容加以限制的法案,使用模糊的语言势必会极大地削弱言论自由。其次,法案作为一部刑法法案,其偏重的刑罚会产生震慑作用,模糊不清的表述也会使民众有可能因惧怕被判刑而放弃自己本来合法的言论。

《儿童在线保护法案》(COPA)

在《通讯规范法案》被判定违宪后,美国议会吸取了法院关于法条表述模糊的意见,重新草拟了《儿童在线保护法案》。

《儿童在线保护法案》从三个方面对管制内容作出了限制。首先,《通讯规范法案》的适用范围包括整个互联网,而《儿童在线保护法案》则限制在万维网上。其次,《儿童在线保护法案》仅适用“商业目的”的通讯。最后,《儿童在线保护法案》仅针对“对未成年人有害”的网络内容加以管制。

《儿童在线保护法案》对内容的限制经受住了《宪法第一修正案》的严格审查。然而,最高法院及地方法院都认为其提供的监管方法仍然过于宽泛(Not the least restrictive),会同时阻挠成年人发表受保护的言论。经过三轮独立的司法诉讼后,《儿童在线保护法案》被判定违宪,在2009年被永久禁止。

《儿童互联网保护法案》(CIPA)

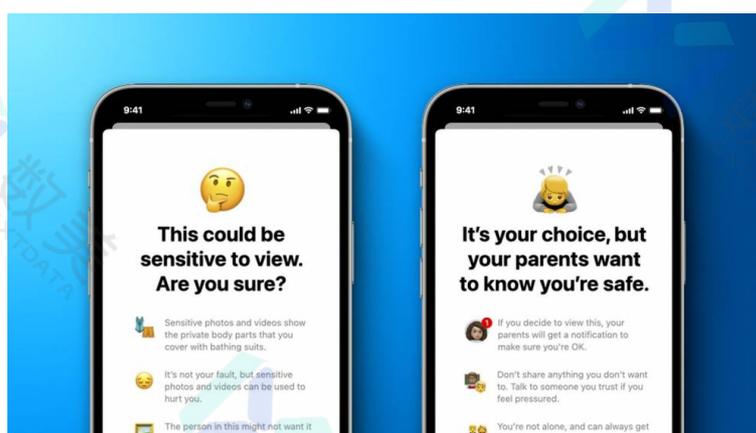
2000年出台的《儿童互联网保护法案》旨在解决儿童有可能通过公共图书馆接触网络色情内容的问题。《儿童互联网保护法案》要求所有接受联邦补贴的公共图书馆在

可上网的电脑上安装过滤软件，拦截有害色情内容。如果符合条件的图书馆拒绝安装过滤软件，它将无法继续享受联邦资金补贴。《儿童互联网保护法案》的出台也引发立法争议，经过几次诉讼后，最高法院最终裁定《儿童互联网保护法案》不会侵犯《宪法第一修正案》，因为图书馆完全可以通过拒绝联邦补贴资助而放弃安装过滤软件。《儿童互联网保护法案》也因此成为美国目前唯一一部有权对互联网内容进行管制的法案。

《儿童在线隐私保护法案》(COPPA)

与前面三部对互联网内容进行管制的法案不同，《儿童在线隐私保护法案》主要针对的是 13 岁以下儿童的隐私保护。COPPA 于 2000 年 4 月开始正式实行，它要求包括线上网站和在线服务机构在内的网络从业者在收集 13 岁以下儿童的个人信息数据之前要通知儿童父母，并征得其明确同意后方可进行信息收集。

从《通讯规范法案》到《儿童在线隐私保护法案》，美国未成年人网络保护立法历经几次废止，立法重点也经历了几次转变。截至目前，儿童隐私保护已成为美国未成年人网络保护司法上贯彻得最为彻底的一个领域，而对未成年人不宜内容的立法管制仍受到较大局限，主要依靠平台公约进行自我约束。例如，苹果在 2021 年 8 月宣布，在其 Messages 应用中更新了一项通信安全功能，当孩子收到色情图片时，图片会变得模糊，并且 Messages 应用程序会显示一条警告，告知孩子可以不查看照片，并提供资源以联系他们信任的人寻求帮助⁸。



儿童用户在收到色情图片时，苹果系统会发送警告

来源：TaiWan News

⁸ 2021.11.10，来源：新浪数码，iOS 15.2 第二个测试版上线新功能：数字遗产、反跟踪和儿童保护，

<https://finance.sina.com.cn/tech/2021-11-10/doc-iktzscyy4642261.shtml>

3) 色情内容

在美国，网络色情是一个非常具有争议的话题。自 20 世纪 90 年代以来，公众、立法者以及法院就如何控制网络色情内容争论不休。反对者希望从道德的角度出发对网络色情内容进行管制，支持者则认为网络色情内容同样受《宪法第一修正案》保护，是公民言论自由的一部分。

单从立法层面而言，在美国通过网络浏览色情内容并不违法。虽然美国最高法院在 1957 年裁定，淫秽内容（Obscenity）不具备“任何救赎社会重要性（Utterly without redeeming social importance）”，因此不受宪法修正案保护。但由于美国司法对言论自由的高度保护，判断一份材料是否“淫秽”的标准也更为严格。

但要注意的是，虽然美国立法对于网络色情内容普遍较为容忍，儿童色情是坚决不能触碰的红线。1990 年，最高法院裁定儿童色情内容不受宪法修正案保护，任何浏览、持有、下载、传播、购买儿童色情内容的行为都是违法的。2016 年 3 月，美剧《吉列合唱团》（Glee）主演 Mark Salling 被警方从其电脑中搜出五万张儿童色情图片，其中包括 10 岁以下的儿童图片，法院以“接收并持有儿童色情物品罪”对其起诉。2021 年，前美国海军使馆安东尼·加百列·欧提兹被发现其通过通讯应用程序共享儿童色情图片，之后因“持有、传播儿童色情制品”被起诉，并被拒绝保释拘留候审。

4) 恐怖主义



克莱斯特彻奇恐怖袭击直播画面

图片来源：中央电视台

随着社交媒体的蓬勃发展，越来越多的恐怖分子和暴力极端分子也开始利用社交媒体开展宣传、招募、运营等。由于绝大多数活跃的互联网平台都是美国企业，平台也日益受到来自政府、股东、公民社会团体、媒体、民众等多方压力，要求平台采取措施对有害内容进行管制。

近年来，多方呼吁美国政府通过立法直接对网络恐怖主义内容进行监管，但从实际情况来看，美国政府的动作还是较为克制，社交平台自身成为网络恐怖主义内容管控的中坚力量。原因在于，首先社交平台比政府更擅长、更懂如何管控平台上的内容。其次，网络上的恐怖主义内容往往是国际问题，而不受单个政府或监管机构的管辖，这就导致非美国企业很可能不会配合美国政府采取行动。再次，如果政府通过罚款、刑罚、反垄断、取消第三方平台的豁免资格等手段要求社交平台采取措施，虽然可能会达到一定的效果，但政府在社交平台企业方的公信力会受到较大折损，企业有可能会拒绝与政府合作。如果立法，政府必须对“恐怖主义内容”等表述进行定义，模糊不清的表述则会催生出大批的法律诉讼，而即使是再清楚的表述，也无法将所有的“恐怖主义内容”囊括在内。

经过长时间的讨论与尝试，美国现在已形成一套政府与行业共同参与的复杂的私下调解机制。美国政府自身不直接介入平台上的内容管制，而主要将这一责任“外包”给企业。一方面，美国政府虽然不会直接制定法律来监管网络上的恐怖主义内容，但它会通过立法权向社交媒体企业施压。另一方面，美国政府也通过提供专家指导、公私合作的方式，鼓励企业采取审核手段对抗平台上的恐怖主义内容。

在公众和美国政府的施压之下，美国各大社交媒体平台开始加快反恐步伐。事实上，美国互联网巨头最先感受到的压力来自欧洲政府。在 2015 年和 2016 年欧洲发生多次恐怖袭击事件后，Facebook、Twitter 和 Google 为回应欧洲各国政府的施压，开始移除其平台上的恐怖主义内容。

随着越来越多的互联网巨头加入这一行列，打击平台恐怖主义内容从最开始的单兵作战变成了行业合作。Facebook、Twitter、Google 以及 Microsoft 共同建立了 Hash 内部共享数据库，里面存储着各自平台标注为恐怖主义的图像及视频 hash。

在 Hash 内部共享数据库的基础上，Facebook、Twitter 和 YouTube 在 2017 年 6 月 26 日宣布组建全球网络反恐论坛（Global Internet Forum to Counter Terrorism，

GIFCT)，旨在对其服务中出现的恐怖主义或暴力极端主义采取强硬措施。

在创立之初，GIFCT 还只是一个行业论坛，由四家企业轮值主席。2019 年，GIFCT 成为一家有着自己的 CEO、运营董事会及独立顾问委员会的独立实体，成员增加至 19 家互联网企业。

2019 年，在新西兰克莱斯特彻奇清真寺发生枪击事件造成 51 人死亡后，多国政府与互联网企业签署“克莱斯特彻奇倡议”，呼吁科技公司和政府共同努力，主动打击网络上的恐怖主义内容。行业自律与企业联合行动成为美国治理网络恐怖主义的中坚力量。



GIFCT
Global Internet Forum
to Counter Terrorism



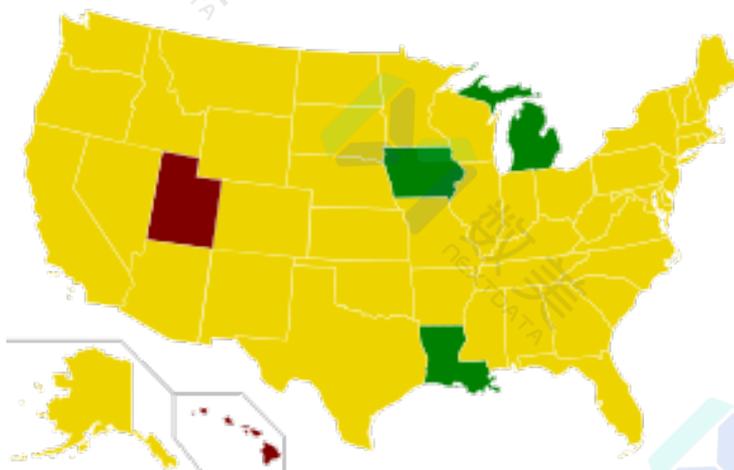
GIFCT 组织

来源：Sirf News

5) 网络赌博

在过去很长一段时间里，美国联邦政府一直将网络赌博列为非法行为。2006 年，美国通过了《取缔非法互联网赌博法案》（Unlawful Internet Gambling Enforcement Act, UIGEA），要求美国的银行和信用卡公司不得为互联网博彩网站提供信用卡、支票、电子支付等转账支付，违者将受处罚。

转变发生在 2011 年，联邦裁定只有体彩赌博在网络上是被禁止的，其他形式的网络赌博不受法律约束，对网络赌博的立法权被交到了各州政府手上。截至 2021 年，美国共有 48 个州宣布网络赌博合法，另外两个仍然禁止网络赌博的州是夏威夷和犹他州。由于犹他州有大量的摩门教徒，受宗教影响，任何形式的赌博在该州都是被禁止的。夏威夷则是因为当地居民认为网络赌博会影响家庭关系和氛围，不支持网络赌博合法化。



美国可以合法赌博的州

(红色：禁止所有类型赌博；绿色：禁止部分类型的赌博；黄色：允许绝大部分类型的赌博)

来源：NCPR

虽然法律并不禁止赌博，但各社交平台都对线上赌博广告做出了限制。以谷歌为例，谷歌仅允许国营实体投放彩票广告，仅允许在在线赌博合法的州投放赛马、体育博彩、在线赌场的广告。同时，广告客户不得定位到未满 21 周岁的用户或获得许可的州以外的用户，必须在登录页或广告素材中添加针对成瘾性和强迫性赌博的危险警告和相关帮助信息。

6) 仇恨言论

管制网络仇恨言论最先是由欧盟提出。2015 年，欧洲难民危机一再升级，网络仇恨言论开始成为欧盟立法机构与网络巨头争论的焦点。2016 年 5 月，欧盟委员会宣布与 Facebook、Twitter、YouTube 和 Microsoft 就控制网络仇恨言论开展合作，四家公司承诺“接到举报后 24 小时内屏蔽和删除相关仇恨言论”。欧盟各国也相继立法对网络仇恨内容进行管制。

美国网络上的仇恨言论也在 2016 年特朗普上台后变得更为严重。但与欧盟通过立法管控仇恨言论不同，由于美国宪法对言论自由的高度保护，再加上《通信端正法》对互联网企业的豁免条款(第 230 条规定，互联网平台无需为其平台上的任何内容负责)，美国政府以技术手段或立法过滤网络内容的情况较少出现。

但这并不意味着美国社会对网络仇恨言论毫无管制。企业自律与行业合作成为互联网公司对抗仇恨言论的主要途径。以 Facebook 为例，Facebook 在其社区守则中将“仇

恨言论”定义为“针对他们受保护的特征，而非观念或习俗直接发起的直接言论攻击，这些特征包括：民族、种族、原国籍、残疾、宗教信仰、种姓、性取向、性别、性别认同，以及严重疾病”，并明确禁止在平台上发布仇恨言论。

2.3.3 加拿大

1) 监管机构

加拿大广播电视和电信委员会（Canadian Radio-television and Telecommunications Commission, CRTC）。是随着《广播法》的出台而专门设立的政府机构，它成立于 1968 年，负责监管加拿大境内的整个广播业，有权限制广播公司的外资比例以及广播内容。上世纪九十年代互联网强势崛起，加拿大政府为鼓励国内互联网和电子商务的发展，决定不为互联网行业设立过多监管限制，鼓励互联网行业通过市场化手段进行自我监管。在 1999 年，CRTC 更是发布《新媒体豁免令》，承诺不干涉互联网媒体服务的发展。这是 CRTC 就互联网媒体服务发出的首个行政令，它的出台直接导致此后二十余年互联网行业不受政府监管。

步入二十一世纪，随着互联网乱象频发，加拿大民众对于管制互联网内容的呼声也越来越高，CRTC 也与立法机关展开多轮辩论，探讨政府是否应当管制互联网内容。近两年，加拿大政府也有过几次意图管制互联网内容的立法尝试，如 2020 年的 Bill C-10 提案和 2021 年的 Bill C-11 提案。但这两部提案更关注通过互联网开展广播业务的音视频媒体，如 Netflix、Disney+ 等流媒体，旨在鼓励网络广播媒体制作更多有关加拿大当地的内容，社交媒体平台仍不受 CRTC 管辖。

截至 2021 年，加拿大尚未设立专门的网络内容监管机构，但这并不意味着在加拿大网络空间是法外地带。任何违反加拿大法律的线下犯罪行为转移到线上，仍然是非法的，如儿童色情、诱骗儿童性行为、仇恨宣传、淫秽内容、线上骚扰、网络威胁等。而网络上其他有争议性的言论，如暴力、针对未成年的烟酒广告、合法色情内容、其他对特定宗教、社会、文化等群体较为敏感的内容等，目前则是依靠行业自律进行管理，政府并未在其中扮演监管角色，社交媒体在现行法律框架下也无法律意义上的审核义务。

2) 立法新趋势

2021年，加拿大政府公布了一项立法提案，计划设立一套网络内容监管机制，迫使社交媒体对其平台上的五种有害内容承担更多责任，即仇恨言论、儿童性剥削、未经同意擅自分享亲密影像、煽动暴力以及恐怖主义。

提案要求平台采取措施撤下以上五种有害内容，包括使用 AI 主动搜索、删除有害内容，24 小时内撤下举报的有害内容等。提案还要求企业公开他们的审核标准、审核记录等操作细则，以接受公众和政府的监督。企业在发现一些特定内容时，还应及时通知执法机构，采取合适的预防措施减少传播。

提案还提议设立一个新的加拿大数字安全委员会（Digital Safety Commission of Canada）来承担新法下的监管与执法工作，包括接收用户投诉、审查平台是否合规、罚款等。对不合规的公司，最高可收取全球收入 5% 或 2500 万美元的罚金。

目前，提案还处于立法阶段，最终是否会通过还是未知数。但从提案来看，加拿大政府也如同美国政府一样，有意愿通过立法来管制网络内容。

三、出海企业风控体系搭建

3.1 人工审核与智能审核

对于多数音视频社交企业，建立内容人工审核团队是一个合乎直觉的选择。实际上许多公司也都拥有人数较多的人审团队，以此来解决当前国内互联网社交平台海量数据的审核需求。随着机器学习和 AI 能力的不断发展，音视频社交平台也逐步选择“机审+人审”的新审核方式。

而对于出海的音视频社交企业，相比起纯人审，“机审+人审”成为了一种必要性选择。除了机审本身速度快、扩容简单、时效性强的优点外，还有以下几类因素：

1) 实时性

在北美、东南亚、中东三大出海地区中，仅有东南亚地区与国内时区接近，其余两地区均有较大时差。如果选择自建内容风控或采用纯人审的方式进行内容审核，很容易造成内容产生高峰与工作时间段的不匹配。例如沙特阿拉伯处于东三区，与国内有 5 小时时差，如果沙特地区的社交平台使用高峰发生在 19 点-23 点，那么国内审核压力最大的时段将是午夜 0 点到凌晨 4 点，如果使用纯人审将直接增加工作量和人力成本。

2) 数据隐私

出海地区往往具有不同的数据隐私规定，规定严苛的区域会要求出海企业保持数据的本地化储存和处理。如若出海企业选择自建内容审核能力，有可能涉嫌跨境数据传输从而违规，但采用“机审+人审”的方式则能够通过本地集群处理大批量数据，再通过当地小规模人审团队进行复核即可。

3) 成本问题

出海企业采用“机审+人审”的另一大优势则在于良好的成本控制能力。在三大出海地区中，广泛使用的语种有英语和阿拉伯语、泰语等，除英语外均属于小语种。如果使用人审的方式对文本和音频进行审核，则需要专业的小语种转译人士或大规模的当地团队，但机审可以对大量正常数据进行过滤，从而减少大批的人审成本。

出海音视频社交企业在构筑风控体系时，主要面临的内容形态有头像、昵称、简介、弹幕、发言、评论、广场内容、直播、语音等，媒体形式为视频、音频、图片、文字四

种，而视频又可拆分为画面截帧和音频，因此风控体系主要的识别能力即为视觉识别、音频识别、文本识别三种。

3.2 视觉识别

3.2.1 视觉识别思路阐述

对于出海企业，尤其是音视频行业出海企业，视觉识别能力已经成为必不可少的云服务基础建设。一方面，出海企业要面临出海国家各具特色的监管要求，需要通过视觉识别建立足够强大的风控系统；另一方面，视觉识别也是实现智能推荐、检索匹配的必要基础。

“视觉识别”是一个宏大的技术主题，从中可以衍生出算法研发、模型训练、策略调优等等研究方向，每个方向都值得展开细致的技术探讨。对于出海企业来说，技术方面可以通过第三方技术服务采购来弥补，更多的精力理应放在自身需求的判断上，关注视觉识别应该解决哪些识别目标的宏观问题上，即“机器面对未知图片，应该具备哪些组成部分的识别能力”的问题，着重分析出海企业在建立视觉识别能力时应考虑哪几方面的识别功能。

对于出海企业，在建立视觉识别能力的过程中可以考虑采用将图片拆解为“人像”、“物品”、“符号标识”、“场景”四大部分的设计思路，从而获得较为完整的画面信息和识别结果。在每个分类中根据自身需要进行细分门类的扩展和下探，或根据出海地区文化禁忌和法律法规进行增减调整。

当然，想要获得更加完整的视觉信息，同样还需要 OCR 技术和 NLP 技术的辅助，甚至要获取视频文件/视频流的音轨内容，才能够更加综合地判定风险的存在。

3.2.2 视觉识别问题分类

1) 人像的相关识别

人像是讨论视觉识别时，最符合直觉的视觉主体，因此对图片中人像的识别能力是解决合规问题和实现更多高级功能的基础。人像包含姿态、肤色、装扮等多种因素，要求出海企业具备较为全面的识别能力。

性别识别：由于海外部分地区的文化环境差异，在对人像进行性别识别时，除了男

性、女性，也要注意存在的部分少数群体。以免直接凭借外在男性/女性特征进行识别产生的偏差，引起不必要的声誉风险。



美国性转空军中校 Bryan Fram

图片来源：WSIU International & National News

体态识别：识别人像姿态和身体形态，如半身像/全身像、站姿/坐姿/卧姿、高矮胖瘦等。因为姿势隐含了基础的倾向信息，例如：半身像更有可能是自拍、卧姿更有可能含有色情风险等等。因此，姿态识别的能力可以作为综合判定图片风险的参考因素之一。而身体形态的高矮胖瘦既能够辅助其他识别的准确性，也能够作为智能推荐的参考因素。

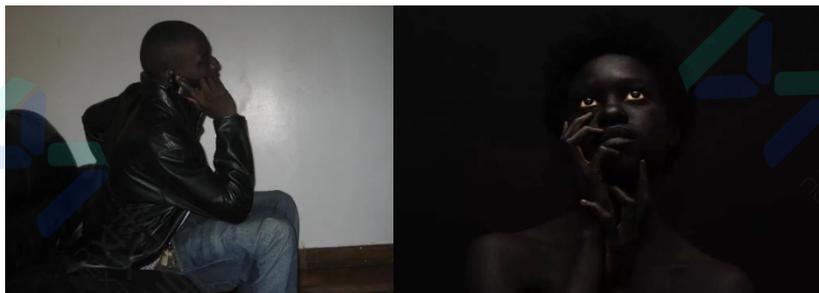
着装识别：对人像着装的识别是风险判定和进阶功能的必备识别能力。在合规方面，着装覆盖皮肤的面积可以作为色情风险的判断因素；着装是否为警服、军装、皇家贵族服饰或者武装分子着装都可作为是否违规的参考因素。当具备着装风格的识别能力时，例如洛丽塔服饰、西装绅士风等，就可以提供兴趣推荐的进阶功能。



一些特殊着装

图片来源于网络

肤色识别：在肤色识别方面，由于国内企业长期训练的人像相关模型通常面向国内业务，因此模型的训练集一般采用了黄种人的样本，在进行海外人像的识别时，就要考虑肤色差异带来的泛化性变化，尤其是一些以往不考虑模型效果的极端场景，例如黑色人种的夜拍图。因此肤色识别并不是要能判定具体的肤色，而是保证模型对人像识别的良好适应能力。



因场景造成识别困难的部分案例

图片来源于网络

2) 物品的相关识别

物品是图片中包含的一大重要视觉元素，这里的“物品”其实包含了无生命的物品和有生命的动植物等。对物品的识别能力直接关系到风险判定和智能推荐的效果，在这里我们可以将“物品”分为敏感物品、普通物品、动植物三类。

敏感物品：敏感物品的识别直接关系到风险判定，画面中出现的某些物体就直接预示了风险，例如枪支弹药、毒品和相关植物、赌桌和老虎机等。但同时，在策略设置上也要考虑地区差异。



部分风险物品示例

图片来源于网络

比如：当枪支出现在北美地区的社交媒体时，单从图片判定是否有风险是模棱两可的，但在中东的禁枪国家就可以直接判定为严重的违禁。同理，像十字架、圣经等具有宗教色彩的物品，在不同国家也具有不同的风险尺度。

普通物品：这里定义的普通物品即各类常见物品，对普通物品的识别更多的服务于智能推荐功能。例如，某位用户的动态图片中出现了多台智能手机、平板电脑、智能手表和电脑设备，那么他很有可能是数码爱好者，或是从事相关产品的产研销工作。



普通物品示例
图片来源于网络

动植物：动植物的识别同样关系到内容合规与智能推荐两方面需求。在内容合规方面，某些动植物因其珍惜性需要识别，比如虎类等保护动物。还有些动植物是因为地区性文化禁忌而需要识别，比如印度的牛、泰国的象和中东地区的猪，都需要针对性地训练相应的动物识别模型。相比起来，智能推荐需求就更加容易理解，例如如果动态图片中频繁出现树木，那么他可能喜欢户外徒步，或者以护林员为工作。至于图片中出现的猫、狗等，也可以用作猜测宠物喜好的依据。



印度的牛崇拜和泰国的象崇拜
图片来源于网络

2) 符号/标识的相关识别

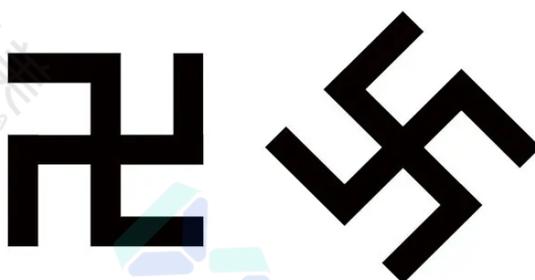
在图片的组成部分中，符号与标识往往是占据较小图像面积的固定图形，但又具备非常明显和强烈的象征意义，在符号和标识中包括了各类国徽会徽、军队标识、Logo、台标、宗教符号等，甚至地图这种形状是否完整具备重大意义的图形也可以视作符号标识的一种。

在这里我们可以分为高敏感标识和低敏感标识。

高敏感标识：这类标识往往因其意义具有相当高的敏感度，在识别时应该采用“宁错勿漏”的灵敏策略。例如：纳粹相关标识、佛教“卍”字标识、十字架标识、地图等。

佛教“卍字符”

纳粹符号



佛教“卍”字符与纳粹“卐”字符

图片来源于网络

这些符号在不同地区也具有不同的监管尺度，比如在大部分地区出现六角星图案不会有什么问题，但中东地区会将其与犹太教联想起来，因此在打入具有国家性宗教或有宗教信仰人口比重很高的国家市场时，对符号和标识的识别要尤为注意。

低敏感标识：这类标识主要囊括了一些常见的其他标识，比如商企 logo、二维码、电视台台标、水印、协会徽标等。对这类标识的识别主要是能够满足平台的个性化需求，比如想判断用户发布的视频是否涉及竞对平台、是否是搬运涉嫌版权侵权，就需要识别画面中是否有 logo、水印、台标等关键标识。

3) 场景的相关识别

在一般画面中，除却人物、物品、标识之后剩余的部分便可视为场景。在某些技术路线的设计中采用目标检测和判定主体风险的思路，但画面背景实际也蕴含了相当丰富的参考信息，同样可以作为视觉识别结果的参考因素。



时尚街拍
图片来源于网络

例如户外、街景、卧室、卫生间等场景已经包含了主体行为的倾向，比如“街景”（场景）+“全身像”（体态）+“潮流”（着装）+“潮流品牌”（logo）的组合就表明图片很有可能是关于时尚街拍的内容，结合主体行为和背景信息往往能形成更加准确的综合判定结果。

3.3 文本识别

3.3.1 文本识别思路阐述

文本识别，一方面是出于监管需要，识别涉政、色情、违禁等敏感内容，一方面是处于维护平台氛围和用户体验需要，识别辱骂、广告、诈骗内容，具体类别如下：

- 涉政文本识别：各类场景中涉政人物、政治事件、宗教、反动分裂，以及恐怖主义等违规文本
- 色情文本识别：淫秽、污秽、色诱、文爱等涉黄内容，支持重度色情、轻度色情分级
- 宗教内容识别：传教、宗教极端言论等内容
- 辱骂文本识别：各类场景中含有污辱、谩骂、诋毁等辱骂内容
- 违禁文本识别：赌博、刀枪、毒品、造假、贩假等违规内容
- 广告导流识别：违反欺诈广告、导流广告；识别利用微信号、手机号、QQ、二维码等开展违法垃圾广告内容；

- 网络诈骗识别：杀猪盘（情感诈骗）；杀鸟盘（兼职刷单诈骗）

在音视频社交应用中，文本场景通常分为公开静态、公开动态和私域场景里。

- 公开静态：个人昵称、签名、介绍等
- 公开动态：动态广场、评论、弹幕
- 私域：私聊、多人聊天

虽然语言不同，文化不一，但整体识别思路和中文语境思路一样，根据出海当地的文化和政策进行调整。文本识别的初级阶段，是通过建立关键词库对产生的文本进行匹配，并使用黑白名单、放行等策略进行审核。进一步则是针对语义的理解、变形的识别和情感的分析。这里以中文为例，说明具体的实现方式。

3.3.2 文本识别问题分类

1) 语义理解的相关识别

智能文本过滤基于海量文本语料训练 NLP 模型，采用 Word2vec 词向量、fastText 文本分类等技术，

Word2vec 是从大量文本语料中以无监督的方式学习语义知识的一种模型，它被大量地用在自然语言处理(NLP)中。通过学习文本来用词向量的方式表征词的语义信息，即通过一个嵌入空间使得语义上相似的单词在该空间内距离很近。

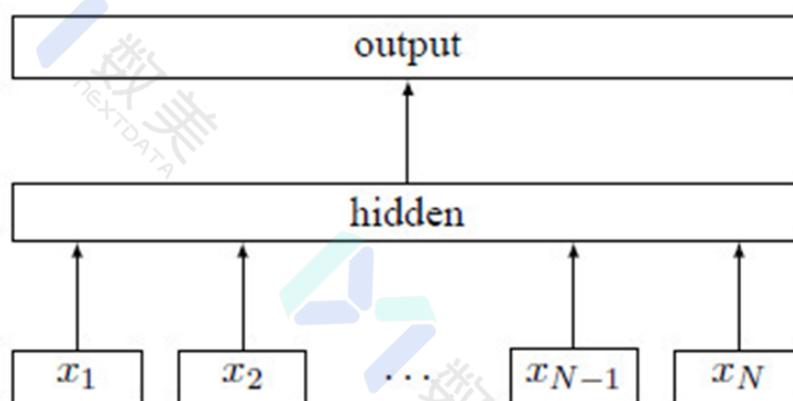
理解人们交谈时的内容，或者敲打文章时的内容——这对于人工智能研究者来说是最大的技术挑战之一，但这也是一个关键需求。自动文本处理在日常计算机使用中相当关键，在网页搜索和内容排名以及垃圾内容分类中占重要组成部分。且当它运行的时候你完全感受不到它。随着在线数据总量越来越大，需要有更灵活的工具来更好的理解这些大型数据集，来提供更加精准的分类结果。

fastText 是一个资料库，能针对文本表达和分类帮助建立量化的解决方案，**fastText** 结合了自然语言处理和机器学习中最成功的理念，包括了使用词袋以及 n-gram 袋表征语句，还有使用子字(**Subword**)信息，并通过隐藏表征在类别间共享信息。叠加 **softmax** 层级(利用了类别不均衡分布的优势)，可以加速运算过程。这些不同概念被用于两个不

同任务：有效文本分类和学习词向量表征。

举例来说：**fastText** 能够学会“男孩”、“女孩”、“男人”、“女人”指代的是特定的性别，并且能够将这些数值存在相关文档中。然后，当某个程序在提出一个用户请求（假设是“我女友现在在儿？”），它能够马上在 **fastText** 生成的文档中进行查找，并理解用户想要问的是有关女性的问题。

除此外，**fastText** 也解决了文本处理领域中，深度神经网络训练以及测试过程缓慢的问题。对于有大量类别的数据集，**fastText** 使用了一个分层分类器（而非扁平式架构）。不同的类别被整合进树形结构中（想象下二叉树而非 list）。



fastText 分层分类器

图片来源于网络

考虑到线性以及多种类别的对数模型，这大大减少了训练复杂性和测试文本分类器的时间。

实验表明，**fastText** 在准确率上与深度学习分类器具有同等水平，特别是在训练和评估速率上要高出几个数量级。使用 **fastText** 能够将训练时间从几天降至几秒，并且在许多标准问题上实现当下最好的表现（例如文本倾向性分析或标签预测）。

文本分类对于商业界来说非常重要。垃圾邮件或钓鱼邮件过滤器可能就是最典型的例子。现在已经有能为一般分类问题（例如 **Vowpal Wabbit** 或 **libSVM**）设计模型的工具，但是 **fastText** 专注于文本分类。这使得在特别大型的数据集上，它能够被快速训练。使用一个标准多核 **CPU**，可以在 10 分钟内训练完超过 10 亿词汇量模型。此外，**fastText** 还能在五分钟内将 50 万个句子分成超过 30 万个类别。

2) 变体识别

大众通常看到的火星文、颜文字、中英数字符号混排等，其实都可以被拆分成【关键词】和字母+数字+字符的【串】，在出海语境下，通常会出现字母+数字+字符的【串】，通过大小写、非英文字符等实现变体。

其中，【微信】二字可以通过汉字音近、汉字形近、拼音音近、语义表达等方式延伸出多种变体，如：

- 可以的哈。这儿发不出去，薇给你发吧
- 可以加个微❤️聊聊不？那个方便点
- WXid_b4iz7Kaegeja22
- 寄拍 📸vx
- 绿色的社交软件
- 可➕VX 13979993427
- 你好亲，XXX 发不了文件，+📧x 免费发给你哦~[萌萌哒R]

【串】可以通过中文形近、中文音近、数字形近、语义干扰、结构干扰等延伸出多种变体，如：

- 二零二一三三九七
- 感觉不错，可以深入了解一下吗，壹柒二零把吴壹武一
- ninē
- ①⑥③⑤⑧⑧④②⑦
- 000000005312311000000005312311
- 5 刚哥 3 更好地跟 1 违规 2 个问题 3 嘎嘎 1 各位哥 1 豆果
- 【爱吃火锅】的拼音就是我的微信

【串】会单独出现，也会和【微信】同时出现，通过顺序干扰、语义干扰、结构干扰等方式，就能延伸出更多变体，如：

- 微信哈哈哈哈哈中间间隔要长长长 dsahg123
- 微【dsahg123】信

- 一纸书封四句诗，四类沾衣抚心叹。
- 八公草木气森森，四月芳林何悄悄。
- 五原纯色旧来迟，六月霜寒表外阴。
- 六出奇花已往开，三点两点淡尤好。
- 三两日来连庆会，二升铛内煮山川。

对于初创公司而言，人审是一个选择，准确率高，但成本也极高，自建模型则对技术和样本有一定要求，初创公司或中小体量平台也很难应对黑灰产不断进化的变体表现形式。

以中文互联网语境为例，数美的【广告-联系方式】标签下，覆盖了微信、QQ、公众号等市面上主流的联系方式，单就【微信】一种，就划分出近 30 种变体形式，主流联系方式（微信、qq、手机号、网址、公众号、百度搜索、微博等）变体特征库达上万吨，可以应对常见的黑产变体，而非中文一旦出现新的变体类型，模型也会及时更新和迭代。对应出海场景，常见的【英文+数字+符号】的邮箱、id 等，都可以呈现出多种变体形式，对应 Facebook、Instagram、WhatsApp、telegram、Line 等多种联系方式。

举个文字的简单案例，一个阿拉伯数字 1，也可以表现为①，但 Unicode 为每种语言中的每个字符设定了统一并且唯一的二进制编码，只要数美资料库里的 unicode 足够多，可以及时实现对字符的变体识别。

而其它的诸如语义干扰、结构干扰等变体形式识别，也是通过给算法“喂”足够多样本，提升模型的识别准确率。

3) 文本情感分析的相关识别

文本的情感分析指的是利用自然语言处理和文本挖掘技术，对带有情感色彩的主观性文本进行分析、处理和抽取的过程，本文仍以中文语境举例。

其中包含五个要素：**Entity**（实体），**Aspect**（属性），**Opinion**（观点）、**Holder**（观点持有者）、**Time**（时间），需要回答的就是一个问题：“什么人？在什么时间？对什么东西？哪一个属性？表达了怎样的情感？”表达者和表达时间的影响基本不对最终的评价态度产生影响，所以情感分析主要针对实体、属性和观点。

例句：XX 汽车的智能控制系统非常先进。

XX 汽车：实体

智能控制系统：属性

先进：观点

文本的情感分析任务包括了词级别的情感分析、句子级情感分析以及目标级情感分析。词级别和句级别的分析是一个词或整个句子的情感正负向，文本中具体的实体、属性不做单独区分。如果文本较复杂，或对准确度要求更高，则需要针对文本中具体的评价对象做出判断，即目标级文本情感分析。目前数美已经能实现对文本中 6 个目标（评价对象）的正负向判断。

词级别：

生日——正向

句级别：

XX 课程干货多，内容细致——正向

目标级：

XX 系列的仪表盘一点也不好看车子整体好看。

仪表盘（属性）-负向；车子整体（属性）-正向

我支持中国政府对 X 国政策的指责。

中国政府（属性）-正向；X 国政策（属性）-负向

比如相亲社区，昵称、个人介绍、私聊等场景是常见风险场景，某客户接入数美正负向情感分析模型后，人审成本大幅降低，减少 80%。另一个案例是网文社区，小说文本是核心风险场景，以往通过关键词进行文本过滤，某客户接入正负向情感分析模型后，同样场景下，准确率提升 30 倍。

3.4 音频识别

3.4.1 音频识别思路阐述

广义上的语音识别技术指的是对语音信号进行识别和理解，并将其转换成相应的文本和命令，代表应用场景有语音交互，包括语音助手、智能音箱等。本篇的音频识别则进一步细化了语音识别的应用场景，主要讨论语音识别技术在音频内容审核上的应用。

音视频社交场景下，音频主要会出现三类信息要素：有语义音频、无语义音频以及说话人声纹信息。这三类信息都有可能承载违规内容，识别它们也因此成为音频合规必须具备的能力。

音频识别按流程大致可分为两步：预处理与识别。首先要对已有的音频文件进行预处理。在线语音场景下的音频信息复杂，存在声源远场、低信噪比、房间混响、回声干扰以及多声源信号干扰等问题，会大大降低语音识别的准确率，因此有必要通过预处理减少对音频识别的干扰。处理好的音频文件会进入模型，通过大量的深度计算对音频中的有语义片段和无语义片段进行识别并标注。

3.4.2 音频识别问题分类

1) 有语义的音频风险识别

有语义音频，即音频中所出现的语言是有含义的，演讲、有声书、相声、有歌词的歌曲均为有语义音频。有语义音频承载了大量的语义信息，其中就有可能包含暴力、色情、诈骗等违规内容，因此对音频中的语义内容进行识别是音频审核的重要能力。

识别有语义音频，普遍使用语音识别技术（ASR）将语音转化为文本，再结合 NLP 和敏感词库等文本识别技术，对转写的文本内容进行审核。

有语义音频的识别主要使用的文本识别技术，已在“第三节文本识别”中详细提及，此处不再详细展开。

2) 无语义的音频风险识别

音频中除了有语义内容，有时还会出现一些无意义的声音，它们无法通过 ASR 进行识别转写，统称为无语义音频。无语义音频虽然并不出现语义内容，但在一些情况下，它们的出现也有可能构成违规。

娇喘

在直播、短视频、音频等领域，ASMR、喊麦等表现形式中存在一定的色情信息。涉黄者往往通过提供一些性暗示，如娇喘，来诱导观众们刷礼物消费，随后将打赏粉丝们加入一个忠粉群，在忠粉群里面进一步传播涉黄内容，让平台审核者防不胜防。这类违规声音类型会混杂在正常对话、背景音中，拥有较强的隐蔽性。如果任由其发展，平台很有可能为淫秽音视频的传播推波助澜。在中东、东南亚这些对色情内容管控较严的地区，除了一般的涉黄语言内容，识别娇喘、呻吟这类特殊声音也应成为音视频审核的必备能力之一。

歌曲

在出海场景下，有些国家对国歌与宗教音乐的传唱与传播有着较为严格的规定，对歌曲的识别也应成为音频识别能力的一大要素。

歌曲的歌词虽然含有语义信息，但唱出来的歌词会有语音语调的变化，再加上有背景音乐的干扰，ASR 识别转写的准确率会大打折扣。此外，仅根据歌词内容有时很难直接判断其是否违规。以歪唱国歌为例，国歌的歌词虽然没有变动，但以歪曲、贬损的方式演唱也会构成违规，这类音频自然也无法通过有语义音频的 ASR 和 NLP 来识别风险内容，而需要结合特定的歌曲识别模型和曲库来进行识别。

空播

除了上面两种违规声音，还有一类情况在音视频社交场景中也经常出现——空播。空播严格意义上并不违规，但有部分主播会利用空播来骗取平台直播时长奖励，这不仅影响用户体验，也会占用平台房间资源。

无语义识别不仅要能识别音频里的声音，也要能结合音频中没有声音的部分，综合判断其是否为空播，从而帮助平台优化资源分配和智能推荐功能。

3) 说话人身份的相关识别

除了识别说话内容，说话人身份也会成为音频内容审核的重要参考依据。在部分场景下，对一些特定人物的声音识别也有助于平台管控合规风险，是风控能力的重要一环。

特定人物声纹识别

由于海外各地文化习俗、监管环境的差异，一些特定人物在网络上的传播动向需要

得到严格控制。例如，在中东地区，出现恐怖组织头目的声音时，应结合语义内容判断这段音频是否有煽动暴力的风险；在泰国，如果音频中出现皇室成员的声音，应谨慎判断其内容是否被用于恶搞等不敬行为。对这类特定人物声音的识别往往是通过建立声纹库，将待审核音频的声音与声纹库进行比对来实现。

未成年人身份识别

不论是国内还是国外，未成年人网络保护一直都是政府监管的重中之重。儿童隐私信息收集、针对儿童的个性化推荐、儿童色情内容的识别等，这一切都要求平台准确识别未成年用户。

音频中的未成年人识别可结合有语义内容和音色、音调等无语义内容来共同实现。对音频中的文本信息进行分析，结合未成年人识别模型，可以判断其中是否含有未成年人的习惯用语或容易暴露其年龄的语言内容。提取说话人的声纹特征也可以辅助判断其是否是未成年人。

声音属性识别

随着音视频社交 APP 开始转向精细化运营，音频识别的应用范围进一步扩大。声音属性识别开始成为助力平台精细化运营的一大亮点，它可以对主播的音色进行识别分类，并打上萝莉音、御姐音、大叔音、青年音、老年音等多种音色标签，从而助力精准化推荐功能的实现。

关于数美科技

数美科技成立于 2015 年 6 月，是一家专业的在线业务风控解决方案提供商，致力于解决在线业务中广泛存在的业务风险与内容风险，为企业数字化转型保驾护航，以两大核心产品：天网——全栈式智能业务风控产品、天净——全栈式智能内容风控产品，推动数字风控行业变革。

数美科技独创全栈式数字风控引擎系统，现已实现全球化 AI SaaS 多集群部署，覆盖中国大陆、欧洲、北美、东南亚、印度等十余个国家和地区，日均风控服务达 30 亿次以上。

目前，数美科技积累了工商银行、银联、小红书、爱奇艺、麦当劳、欧莱雅等全球 3000 余家国内外知名企业的服务和客户成功经验，覆盖音视频社交、游戏、银行、新零售、电商、金融等超 15 个行业，被评为企业数字风控行业领军者。

数美科技总部位于北京，并在杭州、上海、深圳、广州设有研发中心和子公司。团队核心成员均来自百度、阿里、腾讯、360、小米等知名互联网企业，拥有 10 余年搜索、安全、语音等互联网在线产品研发经验，以及相关领域百余项国家级技术专利。

公司拥有人工智能研究院、黑产研究院、舆情中心、政策研究中心等多个专家服务团队，为客户提供业务风控、内容生态治理相关的技术咨询、解决方案和专业服务，并携手中国信通院、中国科学技术大学等权威机构和院校，成立专项技术实验室，实现产学研用一体化。

公司当前已完成 D 轮融资，由 CPE、经纬中国、厚朴投资、腾讯、襄禾资本等知名机构联合投资。

公司产品介绍

天网：数美科技结合多年黑产对抗经验打造的一款全栈式智能业务风控产品，旨在为客户解决营销欺诈、交易风控、数据盗爬、欺诈广告等风险问题，降本增效。

天净：数美科技结合人工智能技术打造的一款全栈式智能内容风控产品，旨在为客户提供一站式的内容风控方案，帮助客户识别文本、图片、音频、视频、网页中出现的涉政、违禁、暴恐、色情、导流广告等内容，规避内容风险，提升运营效率。



数美数字风控研究院

公众号 ID: [ishumei2015](#)



数美科技

公众号 ID: [ShumeiTech](#)

联系我们

电话：400-610-3866

邮箱：pr@ishumei.com

官网：www.ishumei.com