Snap Inc.
加州服务条款报告

2023 年 7 月 1 日至 9 月 30 日

重新提交日期：2024 年 5 月 7 日

草案——律师-委托人特权与保密信息

加州服务条款报告（2023 年 7 月 1 日至 9 月 30 日）（重新提交）
Snap Inc.

重新提交原因

根据《加利福尼亚州商业与职业法典》第 22677 条，Snap Inc.（“Snap”）特此向加利福尼亚州总检察长提交本服务条款报告。这是 Snap 首份《加州服务条款报告》的重新提交版本，覆盖时间为 2023 年 7 月 1 日至 2023 年 9 月 30 日（即 2023 年 第三季度），旨在澄清两处无意的遗漏。首先，本报告已作更新，以反映在相关报告期内，Snap 已在其《社群指南》中制定了禁止外国政治干预的相关政策。其次，本报告已作更新，将“儿童性剥削”列为单独且明确的违规类别。这一变更导致部分数据随之更新，相关内容也已反映在此次重新提交的报告中。Snap 于 2024 年 4 月 1 日提交的《2023 年第四季度服务条款报告》中，已反映了“儿童性剥削”这一新增类别。

我们的条款（《加利福尼亚州商业与职业法典》第 22677 条第 (a) 款第 (1) 项以及第 (4) 款第 (E) 项）

我们致力于在 Snapchat 上营造一个安全、有趣的创作与表达环境。 所有 Snapchat 用户必须遵守我们的服务条款，包括我们的社群指南（统称“**条款**”）。

关于我们如何审核内容以及执行政策的更多背景信息，请参阅我们的《社群指南详解系列》，该系列内容包括对我们的审核、执行和申诉政策政策的说明，以及针对《社群指南》所禁止的各类内容和行为的补充信息。

我们还在安全中心提供安全相关的信息和资源，包括如何举报违反条款的行为或与服务相关的其他安全问题。

上述所有文件的英文版附于本报告之后，且在我们提供 Snapchat 服务的所有加州医疗补助计划 (Medi-Cal) 门槛语言中，均可通过我们的网站获取。

内容审核策略及实践（《加利福尼亚州商业与职业法典》第 22677 条 (a) 款第 (3) 项至第 (4) 项）

我们的条款禁止第 22677 条 (a) 款第 (3) 项所提及类别的内容，具体如下：

第 22677 条 (a) 款中提及的内容类别	我们的《社群指南》所禁止的对应内容类别	相关定义与政策，详见我们的《透明度报告术语表》和《社群指南解读系列》
仇恨言论或种族主义	仇恨言论（属于仇恨内容、恐怖主义及暴力极端主义范畴）	基于个人的种族、肤色、种姓、族裔、国籍、宗教、性取向、性别认同、残疾、退伍军人身份、移民身份、社会经济地位、年龄、体重或怀孕状况，贬低、诋毁他人，或煽动针对任何个人及群体的歧视与暴力行为的内容。如需更多信息，请查看我们关于仇恨内容、恐怖主义和暴力极端主义的说明。
极端主义或思想激进化	恐怖主义与暴力极端主义（属于仇恨内容、恐怖主义和暴力极端主义范畴）	宣扬或支持恐怖主义，或者个人和/或团体为推进政治、宗教、社会、种族或环境等意识形态目标而实施的其他暴力、犯罪行为的内容。它包括宣扬或支持任何境外恐怖主义组织或暴力极端主义仇恨团体的内容，以及推动此类组织招募成员或开展暴力极端主义活动的内容。如需更多信息，请查看我们关于仇恨内容、恐怖主义和暴力极端主义的说明。
虚假信息或错误信息	虚假信息（属于有害的虚假或欺骗性信息）	包含会造成危害或带有恶意的虚假及误导性内容，例如否认悲剧事件属实、未经证实的医疗言论、破坏公共事务流程公信力，以及为达成虚假或误导目的而对内容进行篡改。如需更多信息，请查看我们关于有害虚假或欺骗性信息的说明。

骚扰	骚扰与霸凌	指任何可能导致普通人产生情绪困扰的不受欢迎行为，例如言语辱骂、性骚扰或不受欢迎的性关注。此类别还包括未经同意分享或接收亲密图像 (NCII)。如需更多信息，请查看我们关于骚扰和霸凌的说明。
外国政治干预	虚假信息（属于有害的虚假或欺骗性信息）。	关于虚假信息的定义，请参见上文。 冒充是指帐户谎称与他人或品牌有关联。 如需更多信息，请查看我们关于有害虚假或欺骗性信息的说明。
管制物质分销	毒品（属于非法或受管制活动）	指非法药物（包括假药）的分销和使用，以及其他涉及毒品的非法活动。如需更多信息，请查看我们关于非法或受管制活动的说明。

我们的《审核、执行和申诉说明》以及《严重伤害说明》提供了详细信息，包括以下主题：

- 我们如何通过自动化工具和人工审核相结合的方式来管理内容，
- 我们如何回应用户关于涉嫌违反《社群指南》行为的举报，以及
- 我们如何针对违反《社群指南》的内容和用户进行违规处置。

有关违反我们条款的信息（2023 年 7 月 1 日至 9 月 30 日）（《加利福尼亚州商业与职业法典》第 22677 条 (a) 款第 (5) 项）

根据第 22677 条 (a) 款的要求，下文我们提供了 2024 年 1 月 1 日至 6 月 30 日期间，向我们举报或系统自动识别出的违反《社群指南》行为的详细信息。我们首先提供全球数据，随后提供美国数据。 这些数据不仅涉及第 22677 条 (a)(3) 款中提及的违规内容类别，还更广泛地涵盖了我们的《社群指南》中所指出的各类违规行为。[1]

除非另有说明，本节中使用的术语均依据我们的《透明度术语表》进行定义。

[1] 在本报告中，我们将数据细分为：（i）违规内容类别；（ii）内容或帐户被标记的方式（即通过用户举报还是我们的自动化检测工具）；（iii）内容或帐户被执行处理的方式（即通过人工审核员还是自动化工具）。目前我们无法按内容类型（例如：帖子、评论、消息、用户资料）或媒体类型（例如：文本、图片、视频）对数据进行细分，因为截至2023 年第三季度，我们尚未在全球范围及美国境内对此类数据进行统一追踪留存，因而无法提取相关数据用于报告编制。

草案——律师-委托人特权与保密信息

全球数据

违规类别	标记方式	标记的内容或帐户总数[1]	人工审核员执行的内容[2]	通过自动化工具执行处理的内容	由人工审核员执行处理的独立帐户数 [3]	通过自动化工具执行处理的独立帐户数	针对人工审核员执行的帐户锁定[4]所提出的申诉	针对自动化工具执行的帐户锁定所提出的申诉	经申诉后恢复的帐户数[5]（最初由人工审核员锁定）	经申诉后恢复的帐户数（最初由自动化工具锁定）	由人工审核员执行处理的违规观看率 (VVR)[6]	通过自动化工具执行处理的违规观看率 (VVR)	由人工审核员执行处理的独立违规观看者率[7]	通过自动化工具执行处理的独立违规观看者率
仇恨言论	人工举	189,981	45,028	257	39,567	183	206	5	11	0	0.000193%	0.000001%	0.44%	0.002%
	自动检	148	148	0	132	0	0	0	0	0	0.000000%	0.000000%	0.00%	0.000%
恐怖主义和暴力极端主义	人工举	41,399	835	24	751	21	17	0	1	0	0.000005%	0.000000%	0.01%	0.000%
	自动检	11	11	0	11	0	0	0	0	0	0.000000%	0.000000%	0.00%	0.000%
虚假信息	人工举	216,219	460	10	445	9	3	0	0	0	0.000005%	0.000000%	0.01%	0.000%
	自动检	16	16	0	16	0	0	0	0	0	0.000000%	0.000000%	0.00%	0.000%
假冒行为	人工举	213,879	8,040	36	8,002	33	769	0	51	0	0.000002%	0.000000%	0.01%	0.000%
	自动检	5	5	0	5	0	0	0	0	0	0.000000%	0.000000%	0.00%	0.000%
骚扰与霸凌	人工举	4,531,005	505,999	20,239	414,702	11,285	14,546	943	410	13	0.001143%	0.000044%	1.52%	0.051%
	自动检	2,523	2,481	42	2,268	12	78	3	7	0	0.000002%	0.000000%	0.00%	0.000%
毒品	人工举	177,028	115,835	5,010	84,731	4,118	8,331	1,056	231	5	0.000536%	0.000031%	0.75%	0.062%
	自动检	636,008	286,538	158,894	242,067	128,763	73,446	20,420	1,992	103	0.000101%	0.000010%	0.23%	0.028%
威胁和暴力	人工举	401,227	44,172	5,210	34,555	3,648	747	4	35	0	0.000678%	0.000035%	1.08%	0.064%
	自动检	410	323	11	292	6	42	0	0	0	0.000000%	0.000000%	0.00%	0.000%
自我伤害和自杀	人工举	85,339	15,896	56	14,637	33	18	1	5	0	0.000007%	0.000000%	0.01%	0.000%
	自动检	260	252	0	242	0	2	0	0	0	0.000000%	0.000000%	0.00%	0.000%
垃圾信息	人工举	1,254,516	311,954	514,111	269,775	312,043	7,287	128	108	1	0.000858%	0.000106%	1.28%	0.218%
	自动检	50,890	15,636	35,254	14,084	21,633	443	96	7	0	0.000004%	0.000029%	0.01%	0.021%
武器	人工举	48,967	6,129	568	4,831	409	214	45	6	1	0.000035%	0.000001%	0.06%	0.002%
	自动检	123,755	40,106	66,208	32,953	51,275	612	995	25	8	0.000022%	0.000006%	0.06%	0.016%
其他管制物品	人工举	228,900	68,618	4,582	52,689	2,351	3,989	508	111	4	0.000526%	0.000018%	0.87%	0.029%
	自动检	9,967	9,925	42	8,668	21	389	25	27	1	0.000010%	0.000000%	0.03%	0.001%
色情内容	人工举	2,146,825	794,265	398,293	580,110	249,112	60,534	4,233	747	19	0.004442%	0.001858%	3.08%	1.392%
	自动检	397,538	150,421	194,379	98,190	111,567	11,177	1,392	125	10	0.000061%	0.000011%	0.10%	0.019%
儿童性剥削	人工举	389,163	113,454	2,547	96,106	1,949	13,677	68	2,059	11	0.000300%	0.000020%	0.45%	0.017%
	自动检	168,527	78,427	60,312	54,058	44,284	9,124	9,170	745	2,015	0.000002%	0.000000%	0.00%	0.001%

总计	11,314,506	2,614,974	1,466,085	1,920,608	910,767	205,651	39,092	6,703	2,191	0.008932%	0.002172%	5.99%	1.694%

美国数据

违规类别	标记方式	标记的内容或帐户总数[1]	人工审核员执行的内容[2]	通过自动化工具执行处理的内容	由人工审核员执行处理的独立帐户数 [3]	通过自动化工具执行处理的独立帐户数	针对人工审核员执行的帐户锁定[4]所提出的申诉	针对自动化工具执行的帐户锁定所提出的申诉	经申诉后恢复的帐户数[5]（最初由人工审核员锁定）	经申诉后恢复的帐户数（最初由自动化工具锁定）	由人工审核员执行处理的违规观看率 (VVR)[6]	通过自动化工具执行处理的违规观看率 (VVR)	由人工审核员执行处理的独立违规观看者率[7]	通过自动化工具执行处理的独立违规观看者率
仇恨言论	**人工举**	74,256	26,254	184	22,888	127	118	0	7	0	0.0004208%	0.0000048%	1.316%	0.015%
	自动检	86	86	0	79	0	0	0	0	0	0.0000003%	0.00%	0.001%	0.000%
恐怖主义和暴力极端主义	**人工举**	10,901	197	6	190	4	4	0	0	0	0.0000062%	0.0000001%	0.020%	0.000%
	自动检	6	6	0	6	0	0	0	0	0	0.00%	0.00%	0.000%	0.000%
虚假信息	**人工举**	47,421	235	3	223	3	0	0	0	0	0.0000072%	0.0000002%	0.023%	0.001%
	自动检	10	10	0	10	0	0	0	0	0	0.00%	0.00%	0.000%	0.000%
假冒行为	**人工举**	54,948	2,461	13	2,442	11	241	0	16	0	0.0000001%	0.00%	0.000%	0.000%
	自动检	2	2	0	2	0	0	0	0	0	0.00%	0.00%	0.000%	0.000%
骚扰与霸凌	**人工举报**	1,134,660	166,787	4,658	140,939	3,385	3,987	89	173	9	0.0017937%	0.0000227%	4.261%	0.051%
	自动检	1,189	1,186	3	1,092	2	28	1	4	0	0.0000043%	0.00%	0.014%	0.000%
毒品	**人工举**	76,888	54,098	1,922	39,227	1,655	3,439	166	96	0	0.0014146%	0.0000292%	2.821%	0.111%
	自动检	369,835	170,066	117,427	142,887	93,734	37,681	11,458	909	58	0.0002741%	0.0000334%	0.980%	0.141%
威胁和暴力	**人工举**	117,412	16,571	1,432	13,448	1,057	316	0	22	0	0.0006518%	0.0000524%	1.691%	0.134%
	自动检	222	167	10	153	5	26	0	0	0	0.0000007%	0.0000001%	0.002%	0.000%
自残与自杀	**人工举**	29,226	8,027	8	7,583	8	6	0	3	0	0.0000126%	0.00%	0.040%	0.000%
	自动检	159	153	0	146	0	0	0	0	0	0.00%	0.00%	0.000%	0.000%
垃圾信息	**人工举**	580,657	137,514	360,649	124,895	223,162	1,997	19	22	0	0.0009065%	0.0000995%	2.147%	0.326%
	自动检	15,974	6,304	9,670	6,126	6,276	122	1	2	0	0.0000037%	0.0000129%	0.016%	0.030%
武器	**人工举**	17,212	1,742	80	1,604	72	66	9	3	0	0.0000382%	0.0000009%	0.142%	0.004%
	自动检	99,084	32,206	56,158	26,788	43,961	449	209	17	4	0.0000886%	0.0000241%	0.345%	0.101%
其他管制物品	**人工举**	73,261	13,629	306	11,770	210	340	20	23	1	0.0004930%	0.0000038%	1.482%	0.012%
	自动检	3,539	3,534	5	3,173	3	63	0	10	0	0.0000098%	0.00%	0.041%	0.000%
	人工举	584,728	221,552	127,380	163,531	84,979	16,924	824	233	8	0.0057545%	0.0025898%	8.864%	4.121%

色情内容	**自动检**	109,214	39,790	44,859	27,328	29,015	3,926	238	38	5	0.0001110%	0.0000145%	0.327%	0.042%
	人工举	109,155	25,071	245	22,045	181	13,677	68	2,059	11	0.0001473%	0.0000049%	0.290%	0.011%
儿童性剥削	**自动检**	33,376	12,754	11,707	9,686	8,503	9,124	9,170	745	2,015	0.0000009%	0.0000001%	0.002%	0.000%
总计		3,543,421	940,402	736,725	725,906	486,592	205,651	39,092	6,703	2,191	0.0121398%	0.0028934%	16.290%	4.772%

(1) 因疑似违反《社群指南》而被标记的内容及帐号总数，包括用户举报的内容以及通过自动化工具检测到的内容。为将这些数据按违规内容类型进行分类拆解，本次统计均采用已处置内容对应的最终违规判定依据。如果内容或帐户被标记但未采取处置措施，我们会将相关指标归因于该内容或帐户被标记时所对应的疑似违规类别。
(2) 在 Snapchat 上被采取处置措施的内容（例如：Snap、故事）数量。“处置措施”指针对某条内容或某个帐户所采取的行动（例如：删除、警告、锁定）。
(3) 在 Snapchat 上被采取处置措施的独立帐户数量。例如，如果某个帐户因各种原因被多次处置（例如，用户因发布虚假信息而被警告，随后因骚扰另一用户而帐户被锁定），根据该指标只记为一个被处置帐户。如上所述，“处置措施”指针对某条内容或某个帐户所采取的行动（例如：删除、警告、锁定）。
(4) 用户仅能针对帐户锁定提交申诉。
(5) 我们仅恢复经审核人员确认属于被错误锁定的帐户。
(6) 违规观看率 (VVR)：VVR 是包含违规内容的故事和 Snap 浏览量的百分比，占 Snapchat 上所有故事和 Snap 浏览量的比例。 例如，如果我们的违规观看率 (VVR) 为 0.03%，这意味着在 Snapchat 上每 10,000 次 Snap 和故事浏览中，有 3 次浏览的内容违反了我们的政策。通过该指标，我们能够了解 Snapchat 上的浏览量中有多少百分比来自违反我们《社群指南》（经用户举报或由平台主动处置）的内容。
(7) 独立违规观看者率是指观看过违规内容的独立观看者数量，占整个报告期内（即 2023 年第三季度）活跃独立用户总数的百分比。 例如，如果我们的独立违规观看者率为 0.03%，这意味着在 Snapchat 上相关期间的每 10,000 名活跃用户中，有 3 名观看者看到了违反我们政策的内容。 这一指标有助于我们了解 Snapchat 上有多大比例的用户遇到过违反《社群指南》的内容（经用户举报或由平台主动处置）。

补充信息

虽然第 22677 条并未对此作出要求，但我们认为提供举报受理的中位处理时间 (TAT) 同样具有价值。我们将受理时间 (TAT) 定义为：从我们的信任与安全团队或自动化工具首次接收举报（通常是在举报提交时，或通过自动化手段检测到潜在违规时）起，到最后一个处置动作的时间戳为止的这段时间。如果发生多轮审核，则最终时间根据最后一次采取的操作计算。基于此，我们在全球范围内处理内容和帐户举报的中位时间约为 6 分钟。

如需了解更多关于 Snap 在安全、隐私和透明度方面做法的信息，请访问我们的隐私与安全中心，以及我们的关于透明度报告页面。