Snap Inc.
加州服務條款報告

2023 年 7 月 1 日 - 9 月 30 日

重新提交日期：2024 年 5 月 7 日

草稿 – A/C 特許和保密

加州服務條款報告 (2023 年 7 月 1 日至 9 月 30 日) (重新提交)
Snap Inc.

重新提交原因

根據《加州商業及專業法》第 22677 條，Snap Inc. (以下簡稱「Snap」) 在此向加州總檢察長提交本服務條款報告。 這是 Snap 重新提交的首份加州服務條款報告，涵蓋 2023 年 7 月 1 日至 2023 年 9 月 30 日 (2023 年第三季) 期間，旨在澄清兩項疏忽遺漏。首先，本報告已更新，以呈現在相關報告期間，Snap 已在其社群規範中實施禁止外國政治干預的政策。第二，本報告已更新，將兒童性剝削列為獨立且明確的違反類別。此變更導致部分資料更新，這些資料亦會呈現在此重新提交內容中。Snap 於 2024 年 4 月 1 日提交的 2023 年第四季服務條款報告，已呈現此額外類別的兒童性剝削。

我們的條款《加州商業及專業法》§§22677(a)(1) 與 (4)(E)

我們努力地為 Snapchat 上的創意與自我表達打造安全又有趣的環境。 所有 Snapchat 用戶必須遵守我們的服務條款，包括我們的社群規範 (統稱「**條款**」)。

有關我們如何審核內容與執行政策的其他資訊，請參閱我們的社群規範說明系列，其中包含我們的審核、執行與申訴政策的描述，以及有關社群規範禁止的每個內容類別的其他資訊。

我們也在安全中心提供安全相關資訊與資源，包括有關如何在服務中檢舉違反條款或其他安全問題的指導。

本報告中所附錄的這些文件均以英語呈現，而我們在 Snapchat 網站上提供所有的 Medi-Cal 資格語言。

內容審核政策與實踐《加州商業及專業法》§§22677(a)(3)-(4))

我們的條款禁止第 22677(a)(3) 節中引用的內容類別，具體如下：

第 22677(a) 節中引用的內容類別	**社群規範禁止的相應內容類別**	**相關定義與政策，請參閱我們的透明度報告詞彙表與社群規範解釋系列。**
仇恨言論或種族主義	仇恨言論 (屬於仇恨內容、恐怖主義與暴力極端主義)	基於種族、膚色、種姓、民族、國籍、宗教、性取向、性別認同、殘疾、退伍軍人身分、移民身分、社會經濟身分、年齡、體重或懷孕狀態，貶低、詆毀或宣揚對個人或群組歧視或暴力的內容。如需更多資訊，請詳閱我們關於仇恨內容、恐怖主義和暴力極端主義的說明。
極端主義或激進化	恐怖主義與暴力極端主義 (屬於仇恨內容、恐怖主義與暴力極端主義)	宣揚或支持個人與/或團體，為實現意識型態目標 (例如政治、宗教、社會、種族或環境性質) 而實施的恐怖主義或其他暴力犯罪行為的內容。包括支持任何國外恐怖組織或暴力極端主義仇恨團體的任何內容，以及宣傳此類組織或暴力極端主義活動人才招募的內容。如需更多資訊，請詳閱我們關於仇恨內容、恐怖主義和暴力極端主義的說明。
虛假或錯誤資訊	錯誤資訊 (屬於有害虛假或欺騙性資訊)	包括會造成傷害或惡意的錯誤或誤導性內容，例如否認悲劇事件存在、未經證實的醫學主張或破壞公民程序的完整性，或出於錯誤或誤導性目的而操縱內容。如需更多資訊，請詳閱我們關於有害虛假或欺騙性資訊的說明。
騷擾	騷擾與霸凌	指任何可能導致一般人困擾的不當行為，例如言語虐待、性騷擾或使人感到不適的性關注。此外，這個類別還包括分享或接收未經同意的私密影像 (NCII)。如需更多資訊，請詳閱我們關於騷擾與霸凌的說明。

外國政治干預	錯誤資訊 (屬於有害虛假或欺騙性資訊)。	如需瞭解我們對錯誤資訊的定義，請參閱上文。 假冒他人出現在某帳戶假裝與另一個人或品牌相關聯便構成假冒。 如需更多資訊，請詳閱我們關於有害有害或欺騙性資訊的說明。
管制藥物分銷	毒品 (屬於非法或受管制活動)	指散布與使用非法藥物 (包括假藥) 以及其他涉及毒品的非法活動。如需更多資訊，請詳閱我們關於非法或受管制活動的說明。

我們的審核、執行與申訴說明與嚴重有害說明提供以下主題的詳細資訊：

- 我們如何透過自動化工具與人工審核內容、
- 我們如何回應用戶檢舉涉嫌違反社群規範的行為，以及
- 我們如何對違反《社群規範》的個別內容與用戶進行處置。

有關違反條款 (2023 年 7 月 1 日至 9 月 30 日) 的資訊《加州商業及專業法》§22677(a)(5))

根據 22677(a) 節，以下將提供在 2023 年 7 月 1 日至 9 月 30 日期間向我們檢舉或系統自動偵測的違反社群規範的詳細資訊。 我們首先提供全球數據，然後提供美國數據。這些數據不僅與 22677(a)(3) 節中引用的違反內容類別有關，更與社群規範中廣泛引用的違規內容有關。[1]

除另有規定外，本節中使用的術語均根據我們的透明度詞彙表定義。

[1]在此報告中，我們將資料分為：(i) 違反內容的類別、(ii) 內容或帳戶的標記方式 (即通過報告或我們的自動化偵測工具)，以及 (iii) 內容或帳戶的執行方式 (即通過人工審核人員或自動化工具)。我們目前無法將資料按內容類型 (例如發文、評論、訊息、用戶檔案) 或媒體類型 (例如文字、影像、影片) 進行分類，因為截至 2023 年第三季，我們的檢舉系統尚未設定為即時檢舉此資料。

草稿 – A/C 特許和保密

全球數據

違反類別	行為警告標記	已標記的內容或帳戶總數[1]	由人工審核人員執行的內容[2]	由自動化工具強制執行的內容	由人工審核人員執行的不重複帳戶[3]	由自動化工具強制執行的不重複帳戶	針對由人工審核員執行的帳戶鎖定所提出的申訴[4]	針對由自動化工具強制執行的帳戶鎖定提出的申訴	申訴後恢復的帳戶[5] (最初由人工審核人員鎖定)	申訴後恢復的帳戶 (最初由自動化工具鎖定)	人工審核人員強制執行之內容的暴力瀏覽率 (VVR)[6]	VVR 內容，由自動化工具強制執行	由人工審核者強制執行之內容下不重複的暴力觀看率[7]	由自動化工具強制執行的內容下不重複的違反觀看率
仇恨言論	人類報	189,981	45,028	257	39,567	183	206	5	11	0	0.000193%	0.00001%	0.44%	0.002%
	自動偵	148	148	0	132	0	0	0	0	0	0.00000%	0.00000%	0.00%	0.000%
恐怖主義與暴力極端主義	人類報	41,399	835	24	751	21	17	0	1	0	0.000005%	0.00000%	0.01%	0.000%
	自動偵	11	11	0	11	0	0	0	0	0	0.00000%	0.00000%	0.00%	0.000%
假訊息	人類報	216,219	460	10	445	9	3	0	0	0	0.000005%	0.00000%	0.01%	0.000%
	自動偵	16	16	0	16	0	0	0	0	0	0.00000%	0.00000%	0.00%	0.000%
假冒他人	人類報	213,879	8,040	36	8,002	33	769	0	51	0	0.000002%	0.00000%	0.01%	0.000%
	自動偵	5	5	0	5	0	0	0	0	0	0.00000%	0.00000%	0.00%	0.000%
騷擾與霸凌	人類報	4,531,005	505,999	20,239	414,702	11,285	14,546	943	410	13	0.001143%	0.000044%	1.52%	0.051%
	自動偵	2,523	2,481	42	2,268	12	78	3	7	0	0.000002%	0.00000%	0.00%	0.000%
毒品	人類報	177,028	115,835	5,010	84,731	4,118	8,331	1,056	231	5	0.000536%	0.000031%	0.75	0.062%
	自動偵	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%
威脅與暴力	人類報	401,227	44,172	5,210	34,555	3,648	747	4	35	0	0.000678%	0.000035%	1.08%	0.064%
	自動偵	410	323	11	292	6	42	0	0	0	0.00000%	0.00000%	0.00%	0.000%
自傷與自殺	人類報	85,339	15,896	56	14,637	33	18	1	5	0	0.00007%	0.00000%	0.01%	0.000%
	自動偵	260	252	0	242	0	2	0	0	0	0.00000%	0.00000%	0.00%	0.000%
垃圾訊息	人類報	1,254,516	311,954	514,111	269,775	312,043	7,287	128	108	1	0.000858%	0.000106%	1.28%	0.218%
	自動偵	50,890	15,636	35,254	14,084	21,633	443	96	7	0	0.000004%	0.000029%	0.01%	0.021%
武器	人類報	48,967	6,129	568	4,831	409	214	45	6	1	0.000035%	0.00001%	0.06%	0.002%
	自動偵	123,755	40,106	66,208	32,953	51,275	612	995	25%	8	0.000022%	0.000006%	0.06%	0.016%
其他受管制物品	人類報	228,900	68,618	4,582	52,689	2,351	3,989	508	111	4	0.000526%	0.000018%	0.87%	0.029%
	自動偵	9,967	9,925	42	8,668	21	389	25%	27	1	0.000010%	0.00000%	0.03%	0.001%
色情內容	人類報	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%
	自動偵	397,538	150,421	194,379	98,190	111,567	11,177	1,392	125	10	0.000061%	0.000011%	0.10%	0.019%
兒童性剝削	人類報	389,163	113,454	2,547	96,106	1,949	13,677	68	2,059	11	0.000300%	0.000020%	0.45%	0.017%
	自動偵	168,527	78,427	60,312	54,058	44,284	9,124	9,170	745	2,015	0.000002%	0.00000%	0.00%	0.001%

總計	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%

美國數據

違反類別	**行為警告標記**	**已標記的內容或帳戶總數[1]**	**由人工審核人員執行的內容[2]**	**由自動化工具強制執行的內容**	**由人工審核人員執行的不重複帳戶[3]**	**由自動化工具強制執行的不重複帳戶**	**針對由人工審核員執行的帳戶鎖定所提出的申訴[4]**	**針對由自動化工具強制執行的帳戶鎖定提出的申訴**	**申訴後恢復的帳戶[5] (最初由人工審核人員鎖定)**	**申訴後恢復的帳戶 (最初由自動化工具鎖定)**	**人工審核人員強制執行之內容的暴力瀏覽率 (VVR))[6]**	**由自動化工具強制的內容的暴力瀏覽率 (VVR)**	**由人工審核者強制執行之內容下不重複的暴力觀看率[7]**	**由自動化工具強制執行的內容下不重複的違反觀看率**
仇恨言論	**人類報**	74,256	26,254	184	22,888	127	118	0	7	0	0.0004208%	0.0000048%	1.316%	0.015%
	自動偵	86	86	0	79	0	0	0	0	0	0.000003%	0.000000%	0.001%	0.000%
恐怖主義與暴力極端主義	**人類報**	10,901	197	6	190	4	4	0	0	0	0.0000062%	0.000001%	0.020%	0.000%
	自動偵	6	6	0	6	0	0	0	0	0	0.000000%	0.000000%	0.000%	0.000%
假訊息	**人類報**	47,421	235	3	223	3	0	0	0	0	0.0000072%	0.000002%	0.023%	0.001%
	自動偵	10	10	0	10	0	0	0	0	0	0.000000%	0.000000%	0.000%	0.000%
假冒他人	**人類報**	54,948	2,461	13	2,442	11	241	0	16	0	0.000001%	0.000000%	0.000%	0.000%
	自動偵	2	2	0	2	0	0	0	0	0	0.000000%	0.000000%	0.000%	0.000%
騷擾與霸凌	**人類報告**	1,134,660	166,787	4,658	140,939	3,385	3,987	89	173	9	0.0017937%	0.0000227%	4.261%	0.051%
	自動偵	1,189	1,186	3	1,092	2	28	1	4	0	0.0000043%	0.000000%	0.014%	0.000%
毒品	**人類報**	76,888	54,098	1,922	39,227	1,655	3,439	166	96	0	0.0014146%	0.0000292%	2.821%	0.111%
	自動偵	369,835	170,066	117,427	142,887	93,734	37,681	11,458	909	58	0.0002741%	0.0000334%	0.980%	0.141%
威脅與暴力	**人類報**	117,412	16,571	1,432	13,448	1,057	316	0	22	0	0.0006518%	0.0000524%	1.691%	0.134%
	自動偵	222	167	10	153	5	26	0	0	0	0.000007%	0.000001%	0.002%	0.000%
自傷與自殺	**人類報**	29,226	8,027	8	7,583	8	6	0	3	0	0.0000126%	0.000000%	0.040%	0.000%
	自動偵	159	153	0	146	0	0	0	0	0	0.000000%	0.000000%	0.000%	0.000%
垃圾訊息	**人類報**	580,657	137,514	360,649	124,895	223,162	1,997	19	22	0	0.0009065%	0.0000995%	2.147%	0.326%
	自動偵	15,974	6,304	9,670	6,126	6,276	122	1	2	0	0.0000037%	0.0000129%	0.016%	0.030%
武器	**人類報**	17,212	1,742	80	1,604	72	66	9	3	0	0.0000382%	0.0000009%	0.142%	0.004%
	自動偵	99,084	32,206	56,158	26,788	43,961	449	209	17	4	0.0000886%	0.0000241%	0.345%	0.101%
其他受管制物品	**人類報**	73,261	13,629	306	11,770	210	340	20	23	1	0.0004930%	0.0000038%	1.482%	0.012%
	自動偵	3,539	3,534	5	3,173	3	63	0	10	0	0.0000098%	0.000000%	0.041%	0.000%
	人類報	584,728	221,552	127,380	163,531	84,979	16,924	824	233	8	0.0057545%	0.0025898%	8.864%	4.121%

色情內容	**自動偵**	109,214	39,790	44,859	27,328	29,015	3,926	238	38	5	0.0001110%	0.0000145%	0.327%	0.042%
	人類報	109,155	25,071	245	22,045	181	13,677	68	2,059	11	0.0001473%	0.0000049%	0.290%	0.011%
兒童性剝削	**自動偵**	33,376	12,754	11,707	9,686	8,503	9,124	9,170	745	2,015	0.0000009%	0.000001%	0.002%	0.000%
總計		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) 因可能違反社群規範的內容或帳戶總數，包括向我們檢舉的內容或帳戶總數，以及透過我們的自動化工具偵測的內容或帳戶總數。為了將這些資料分解為違規內容的類別，我們使用了採取執法行動的最終執法原因。如果內容或帳戶受到標記但未採取執法行動，我們會將指標歸因於受標記內容或帳戶的疑似違反類別。

(2) Snapchat 上已處置內容 (例如 Snap 和故事) 的數量。「處置」是指對內容或帳戶採取的行動 (例如刪除、警告和鎖定)。

(3) Snapchat 上已處置的不重複帳戶數量。例如，如果單一帳戶因各種原因多次被處置 (例如，用戶因發佈虛假資訊而受到警告，隨後因騷擾其他用戶而遭到刪除) ，此指標中僅會計算一個帳戶。如上所述，「處置」是指對內容或帳戶採取的行動 (例如刪除、警告、鎖定) 。

(4) 用戶只能針對帳戶鎖定提交申訴。

(5) 我們只會復原我們的版主認定被錯誤鎖定的帳户。

(6) 違規內容收視率是包含違規內容之故事與 Snap 的觀看次數，佔 Snapchat 上所有故事和 Snapchat 觀看次數的比率。舉例而言，如果 VVR 為 0.03%，表示 Snapchat 上每 10,000 次 Snap 和故事觀看次數，有 3 次違反我們政策的內容。這項指標讓我們瞭解，Snapchat 上的觀看次數，有多少百分比是觀看違反社群規範的內容 (遭到檢舉或被我們強制執行的內容)。

(7) 不重複違規內容收視率是指在整個報告期間 (即 2023 年上半年) 內，不重複活躍用戶中，看到違反內容的不重複觀眾所佔的比例。 舉例而言，如果我們的不重複違規內容收視率為 0.03%，代表在相關期間 Snapchat 上每 10,000 名活躍用戶，就有 3 名觀眾觀看了違反我們政策的內容。此指標讓我們瞭解 Snapchat 上的用戶中有多少比例會違反我們的社群規範 (已檢舉或主動執行) 的內容。

其他資訊

雖然 22677 節未要求，但我們仍認為提供平均處理時間 (TAT) 對於回覆檢舉和申訴相當有用。我們將 TAT 定義為我們的信任與安全團隊或自動化工具首次收到檢舉 (通常透過自動化方式提交或偵測到潛在違規檢舉) 到上次執法行動時間戳之間的時間。如果歷經數次審核，會以最後一次處置的最終時間為主。考量到這一點，我們內容與帳戶報告的全球 TAT 中位數約為 6 分鐘。

如需進一步瞭解 Snap 的安全、隱私與透明度方法，請造訪我們的隱私與安全中心，以及我們的透明度報告頁面。