

Snap Inc.

Отчёт об Условиях оказания услуг в Калифорнии

1 июля — 30 сентября 2023 г.



Отправлен повторно: 7 мая 2024 г.

Отчёт об Условиях оказания услуг в Калифорнии (1 июля — 30 сентября 2023 г.) (повторная подача)
Snap Inc.

Причина повторной подачи

В соответствии со Статьёй 22677 Кодекса о предпринимательской и профессиональной деятельности Калифорнии Snap Inc. (далее — Snap) настоящим представляет генеральному прокурору Калифорнии Отчёт об Условиях оказания услуг. Это повторное представление первого Отчёта Snap об Условиях оказания услуг в Калифорнии за период с 1 июля по 30 сентября 2023 г. (третий квартал 2023 г.), который призван исправить два непреднамеренных упущения. Во-первых, в этот отчёт внесены изменения, отражающие тот факт, что в течение соответствующего отчётного периода Snap применяла правила, запрещающие иностранное политическое вмешательство в соответствии со своими Правилами сообщества. Во-вторых, в данный отчёт внесены изменения, в соответствии с которыми сексуальная эксплуатация детей выделена в отдельную категорию нарушений. Это изменение приводит к изменению определённых данных, что также нашло отражение в данной повторной подаче. В отчёте об Условиях оказания услуг Snap за четвёртый квартал 2023 года, который был представлен 1 апреля 2024 года, уже отражена эта дополнительная категория — «Сексуальная эксплуатация детей».

Наши Условия (Кодекс о предпринимательской и профессиональной деятельности Калифорнии, Статьи 22677(a)(1) и (4)(E))

Мы стремимся обеспечить в Snapchat безопасную, увлекательную среду для творчества и самовыражения. Все пользователи Snapchat обязуются соблюдать наши [Условия оказания услуг](#), включая наши [Правила сообщества](#) (совместно именуемые «Условия»).

Дополнительную информацию о том, как мы модерлируем контент и обеспечиваем соблюдение наших правил, можно найти в нашей серии пояснений к Правилам сообщества, которые включают в себя описание наших правил [модерации, правоприменения и обжалования](#), а также дополнительную информацию о каждой категории контента, запрещённой нашими [Правилами сообщества](#).

Мы также предоставляем информацию и ресурсы по безопасности в нашем [Центре безопасности](#), включая руководство о том, [как пожаловаться на нарушения](#) наших Условий или другие проблемы безопасности в нашем сервисе.

Эти документы прилагаются к настоящему отчёту на английском языке. Они доступны на нашем веб-сайте на всех языках, на которых мы предлагаем Snapchat.

Правила и практика модерации контента (Кодекс о предпринимательской и профессиональной деятельности Калифорнии, Статьи 22677(a)(3)–(4))

Наши Условия запрещают категории контента, указанные в Статье 22677(a)(3), а именно:

Категория контента, указанная в Статье 22677(a)	Соответствующая категория контента, запрещённого нашими Правилами сообщества	Соответствующие определения и правила, представленные в нашем Глоссарии отчёта о прозрачности и серии пояснений к Правилам сообщества
Риторика ненависти или расизм	Риторика ненависти (относится к категориям «Контент, разжигающий ненависть», «Терроризм» и «Насильственный экстремизм»)	Контент, который унижает отдельного человека или группу лиц либо пропагандирует дискриминацию или ненависть в их отношении по признаку расы, цвета кожи, касты, этнической принадлежности, национального происхождения, религии, сексуальной ориентации, гендерной идентичности, инвалидности, статуса ветерана, иммиграционного статуса, социально-экономического статуса, возраста, веса или беременности. Для получения дополнительной информации ознакомьтесь с нашим пояснением к правилам в отношении контента, разжигающего ненависть, терроризма и насильственного экстремизма .

Экстремизм и радикализация	Терроризм и насильственный экстремизм (относится к категориям «Контент, разжигающий ненависть», «Терроризм» и «Насильственный экстремизм»)	Контент, пропагандирующий или поддерживающий терроризм или другие насильственные и преступные действия, совершаемые отдельными лицами и/или группами лиц для достижения идеологических целей, в частности политического, религиозного, социального, расового и экологического характера. Сюда входит любой контент, который поощряет или поддерживает любую иностранную террористическую организацию или экстремистскую группу разжигающую ненависть, а также контент, поощряющий вербовку в подобные организации или насильственные экстремистские группы. Для получения дополнительной информации ознакомьтесь с нашим пояснением к правилам в отношении контента, разжигающего ненависть, терроризма и насильственного экстремизма .
Дезинформация или ложная информация	Ложная информация (относится к категории «Вредоносная ложная или вводящая в заблуждение информация»)	Включает ложный или вводящий в заблуждение контент, который причиняет вред или несёт злой умысел, в том числе отрицание фактов трагических событий, необоснованные утверждения медицинского характера, подрыв доверия к гражданским процессам или манипулирование контентом с целью введения в заблуждение. Для получения дополнительной информации ознакомьтесь с нашим пояснением к правилам в отношении вредоносной ложной или вводящей в заблуждение информации .
Оскорбления	Оскорбления и буллинг	Относится к любому нежелательному поведению, которое может вызвать у обычного человека эмоциональный стресс, например, словесные оскорбления, сексуальные домогательства или нежелательное сексуальное внимание. В эту категорию также входит распространение интимных изображений без согласия (NCII). Для получения дополнительной информации ознакомьтесь с нашим пояснением к правилам в отношении оскорблений и буллинга .
Иностранное политическое вмешательство	Ложная информация (относится к категории «Вредоносная ложная или вводящая в заблуждение информация»).	Наше определение ложной информации см. выше. Самозванство имеет место, когда владелец аккаунта выдаёт себя за другого человека или ложно связывает себя с каким-либо брендом. Для получения дополнительной информации ознакомьтесь с нашим пояснением к правилам в отношении вредоносной ложной или вводящей в заблуждение информации .
Распространение подконтрольных веществ	Наркотические вещества (относится к категории «Незаконная или регулируемая деятельность»)	Относится к распространению и использованию незаконных наркотических веществ (в том числе поддельных таблеток), а также к другой незаконной деятельности, связанной с наркотическими веществами. Для получения дополнительной информации ознакомьтесь с нашим пояснением к правилам в отношении незаконной или регулируемой деятельности .

Наше [пояснение по модерации, правоприменению и обжалованию](#), а также [пояснение к правилам в отношении серьёзного вреда](#) содержат подробную информацию, в частности, о следующем:

- Способы модерации контента с помощью автоматизированных инструментов и проверки модераторами.
- Способы реагирования на сообщения пользователей о предполагаемых нарушениях наших Правил сообщества.
- Способы правоприменения в отношении отдельных единиц контента и пользователей, нарушающих наши Правила сообщества.

Информация о нарушениях наших Условий (1 июля — 30 сентября 2023 года) (Кодекс о предпринимательской и профессиональной деятельности Калифорнии, Статья 22677(а)(5))

Ниже мы приводим подробную информацию о нарушениях наших Правил сообщества, на которые нам пожаловались пользователи или которые были автоматически обнаружены нашими системами в период с 1 июля по 30 сентября 2023 года в соответствии со Статьёй 22677(а). Сначала мы приводим данные по всему миру, а затем — по США. Эти показатели относятся не только к категориям контента, нарушающего правила, которые указаны в Статье 22677(а)(3), но и, в более широком смысле, к нарушениям, упомянутым в наших Правилах сообщества.¹

¹ В этом отчёте мы разбили данные по следующим категориям: (i) типы контента, нарушающего правила; (ii) способы обнаружения контента или аккаунтов (т. е. по жалобе или с помощью наших автоматизированных инструментов); (iii) способ принятия мер в отношении контента или аккаунтов (т. е. с помощью модераторов или автоматизированных инструментов). На данный момент мы не можем разбить данные по типам контента (например, публикации, комментарии, сообщения, профили пользователей) или по типам медиафайлов (например, текст, изображения, видео), поскольку на третий квартал 2023 года наши системы отчётности не были настроены на предоставление таких данных.

ПРОЕКТ — ЗАЩИЩЕНО АДВОКАТСКОЙ ТАЙНОЙ И КОНФИДЕНЦИАЛЬНО

Если не указано иное, термины, используемые в настоящем разделе, определены в соответствии с нашим [Глоссарием отчёта о прозрачности](#).

ПРОЕКТ — ЗАЩИЩЕНО АДВОКАТСКОЙ ТАЙНОЙ И КОНФИДЕНЦИАЛЬНО

Оскорблени я и буллинг	Жалоба пользо вателя	4 531 00 5	505 999	20 239	414 702	11 285	14 546	943	410	13	0,001143%	0,000044 %	1,52%	0,051%
	Автоматическое обнару	2523	2481	42	2268	12	78	3	7	0	0,000002%	0,000000 %	0,00%	0,000%
Наркотичес кие вещества	Жалоба пользо вателя	177 028	115 835	5010	84 731	4118	8331	1056	231	5	0,000536%	0,000031 %	0,75%	0,062%
	Автоматическое обнару	636 008	286 538	158 894	242 067	128 763	73 446	20 420	1992	103	0,000101%	0,000010 %	0,23%	0,028%
Угрозы и насилие	Жалоба пользо вателя	401 227	44 172	5210	34 555	3648	747	4	35	0	0,000678%	0,000035 %	1,08%	0,064%
	Автоматическое обнару	410	323	11	292	6	42	0	0	0	0,000000%	0,000000 %	0,00%	0,000%
Членовреди тельство и самоубийс тва	Жалоба пользо вателя	85 339	15 896	56	14 637	33	18	1	5	0	0,000007%	0,000000 %	0,01%	0,000%
	Автоматическое обнару	260	252	0	242	0	2	0	0	0	0,000000%	0,000000 %	0,00%	0,000%
Спам	Жалоба пользо вателя	1 254 51 6	311 954	514 111	269 775	312 043	7287	128	108	1	0,000858%	0,000106 %	1,28%	0,218%
	Автоматическое обнару	50 890	15 636	35 254	14 084	21 633	443	96	7	0	0,000004%	0,000029 %	0,01%	0,021%
Оружие	Жалоба пользо вателя	48 967	6129	568	4831	409	214	45	6	1	0,000035%	0,000001 %	0,06%	0,002%
	Автоматическое обнару	123 755	40 106	66 208	32 953	51 275	612	995	25	8	0,000022%	0,000006 %	0,06%	0,016%

ПРОЕКТ — ЗАЩИЩЕНО АДВОКАТСКОЙ ТАЙНОЙ И КОНФИДЕНЦИАЛЬНО

Риторика ненависти	Автоматическое обнару	86	86	0	79	0	0	0	0	0	0,0000003%	0,0000000%	0,001%	0,000%
Терроризм и насильственный экстремизм	Жалоба пользователя	10 901	197	6	190	4	4	0	0	0	0,0000062%	0,0000001%	0,020%	0,000%
	Автоматическое обнару	6	6	0	6	0	0	0	0	0	0,0000000%	0,0000000%	0,000%	0,000%
Ложная информация	Жалоба пользователя	47 421	235	3	223	3	0	0	0	0	0,0000072%	0,0000002%	0,023%	0,001%
	Автоматическое обнару	10	10	0	10	0	0	0	0	0	0,0000000%	0,0000000%	0,000%	0,000%
Самозванство	Жалоба пользователя	54 948	2461	13	2442	11	241	0	16	0	0,0000001%	0,0000000%	0,000%	0,000%
	Автоматическое обнару	2	2	0	2	0	0	0	0	0	0,0000000%	0,0000000%	0,000%	0,000%
Оскорбления и буллинг	Жалоба пользователя	1 134 660	166 787	4658	140 939	3385	3987	89	173	9	0,0017937%	0,0000227%	4,261%	0,051%
	Автоматическое обнару	1189	1186	3	1092	2	28	1	4	0	0,0000043%	0,0000000%	0,014%	0,000%
Наркотические вещества	Жалоба пользователя	76 888	54 098	1922	39 227	1655	3439	166	96	0	0,0014146%	0,0000292%	2,821%	0,111%
	Автоматическое обнару	369 835	170 066	117 427	142 887	93 734	37 681	11 458	909	58	0,0002741%	0,0000334%	0,980%	0,141%
	Жалоба пользователя	117 412	16 571	1432	13 448	1057	316	0	22	0	0,0006518%	0,0000524%	1,691%	0,134%

Угрозы и насилие	Автоматическое обнару	222	167	10	153	5	26	0	0	0	0,0000007%	0,0000001%	0,002%	0,000%
	Жалоба пользователя	29 226	8027	8	7583	8	6	0	3	0	0,0000126%	0,0000000%	0,040%	0,000%
Членовредительство и самоубийства	Автоматическое обнару	159	153	0	146	0	0	0	0	0	0,0000000%	0,0000000%	0,000%	0,000%
	Жалоба пользователя	580 657	137 514	360 649	124 895	223 162	1997	19	22	0	0,0009065%	0,0000995%	2,147%	0,326%
Спам	Автоматическое обнару	15 974	6304	9670	6126	6276	122	1	2	0	0,0000037%	0,0000129%	0,016%	0,030%
	Жалоба пользователя	17 212	1742	80	1604	72	66	9	3	0	0,0000382%	0,0000009%	0,142%	0,004%
Оружие	Автоматическое обнару	99 084	32 206	56 158	26 788	43 961	449	209	17	4	0,0000886%	0,0000241%	0,345%	0,101%
	Жалоба пользователя	73 261	13 629	306	11 770	210	340	20	23	1	0,0004930%	0,0000038%	1,482%	0,012%
Другие подконтрольные товары	Автоматическое обнару	3539	3534	5	3173	3	63	0	10	0	0,0000098%	0,0000000%	0,041%	0,000%
	Жалоба пользователя	584 728	221 552	127 380	163 531	84 979	16 924	824	233	8	0,0057545%	0,0025898%	8,864%	4,121%
Контент сексуального характера	Автоматическое обнару	109 214	39 790	44 859	27 328	29 015	3926	238	38	5	0,0001110%	0,0000145%	0,327%	0,042%
	Жалоба пользователя	109 155	25 071	245	22 045	181	13 677	68	2059	11	0,0001473%	0,0000049%	0,290%	0,011%

Сексуальная эксплуатация детей	Автоматическое обнаружение	33 376	12 754	11 707	9686	8503	9124	9170	745	2015	0,0000009%	0,0000001%	0,002%	0,000%
Всего		3 543 421	940 402	736 725	725 906	486 592	205 651	39 092	6703	2191	0,0121398%	0,0028934%	16,290%	4,772%

- (1) Общее количество единиц контента или аккаунтов, которые были отмечены как потенциально нарушающие наши Правила сообщества, включая те, на которые нам пожаловались и которые были обнаружены с помощью наших автоматизированных инструментов. Для разбивки этих данных по категориям контента, нарушающего правила, мы использовали основную причину правоприменительных мер в тех случаях, когда они были осуществлены. Если контент или аккаунт отметили, как нарушающий правила, но не было принято никаких мер, мы относим эти метрики к категории предполагаемого нарушения, по которой контент или аккаунт был помечен.
- (2) Количество единиц контента (например, Snap, историй), в отношении которых в Snapchat были приняты меры. «Правоприменение» относится к мерам, принятым в отношении какого-либо контента или аккаунта (например, удаление, отправка предупреждения, блокировка).
- (3) Количество уникальных аккаунтов, в отношении которых в Snapchat были приняты меры. Например, если к одному аккаунту меры применялись неоднократно и по разным причинам (например, пользователь получил предупреждение за публикацию ложной информации, а затем был удалён за оскорбление другого пользователя), для данной метрики аккаунт учитывается только один раз. Как указано выше, «правоприменение» относится к мерам, принятым в отношении какого-либо контента или аккаунта (например, удаление, отправка предупреждения, блокировка).
- (4) Пользователи могут обжаловать решение только в отношении блокировки аккаунтов.
- (5) Мы восстанавливаем только те аккаунты, которые, по мнению наших модераторов, были заблокированы ошибочно.
- (6) Доля просмотров с нарушениями (VVR) — это показатель просмотров историй и Snap, которые содержат контент, нарушающий правила, в процентах от всех просмотров историй и Snap в Snapchat. Например, если доля просмотров с нарушениями составляет 0,03%, то на каждые 10 000 просмотров Snap и историй в Snapchat приходится 3 просмотра контента, нарушающего наши правила. Эта метрика позволяет нам понять, какой процент просмотров в Snapchat приходится на контент, нарушающий наши Правила сообщества (на который поступили жалобы либо в отношении которого заблаговременно были приняты меры).
- (7) Доля уникальных пользователей, просмотревших материалы с нарушениями, — это число уникальных пользователей, которые увидели контент, нарушающий правила, в процентах от уникальных пользователей, активных на протяжении отчётного периода (т. е. третий квартал 2023 года). Например, если доля уникальных пользователей, просмотревших материалы с нарушениями, составляет 0,03%, это означает, что на каждые 10 000 активных пользователей в течение соответствующего периода в Snapchat приходится 3 пользователя просмотревших контент, нарушающий наши правила. Эта метрика позволяет нам понять, какой процент пользователей Snapchat сталкивается с контентом, нарушающим наши Правила сообщества (на который поступили жалобы либо в отношении которого заблаговременно были приняты меры).

Дополнительная информация

Мы также считаем полезным предоставить наши данные о среднем времени обработки (TAT) при рассмотрении жалоб и апелляций, хотя это не требуется в соответствии со Статьёй 22677. Мы определяем среднее время обработки (TAT) как период времени от момента, когда наши специалисты по вопросам доверия и безопасности или автоматизированные инструменты впервые получают жалобу (обычно при её отправке пользователем или при обнаружении нарушения с помощью автоматизированных средств), до момента последнего действия по правоприменению. Если проверка проводится в несколько этапов, окончательное время рассчитывается по времени принятия последней меры. С учётом этого наш средний показатель времени обработки (TAT) в отношении жалоб на контент и аккаунты по всему миру составляет примерно 6 минут.

Для получения дополнительной информации о подходе Snap к обеспечению безопасности, конфиденциальности и прозрачности посетите наш [Центр конфиденциальности, безопасности и политики](#) и нашу страницу [Об отчётах о прозрачности](#).