

Snap Inc.

Relazione sui Termini di servizio per New York

1° luglio - 30 settembre 2025



Inviata: 1° gennaio 2026

Relazione sui Termini di servizio per New York (1° luglio - 30 settembre 2025)

Snap Inc.

Snap Inc. invia la presente relazione sui Termini di Servizio al Procuratore generale di New York, come previsto dall'Art. 42, §§ 1100-1104. La presente relazione Termini di servizio per New York riguarda il periodo compreso tra il 1° luglio 2025 e il 30 settembre 2025 (3° trimestre 2025).

I nostri Termini (NY Gen. Bus. L. §§ 1101, 1102)

Ci impegniamo a fornire un ambiente sicuro e divertente che incoraggi gli Snapchatter a essere creativi e a esprimersi in libertà. Tutti gli utenti Snapchat devono rispettare i nostri [Termini di servizio](#), incluse le nostre [Linee Guida per la Community](#) (denominati collettivamente "Termini"). Ulteriori dettagli su come moderiamo i contenuti e applichiamo le nostre politiche sono disponibili nella nostra serie esplicativa sulle Linee Guida per la Community, che include una descrizione delle nostre politiche di [moderazione, applicazione delle norme e ricorsi](#), e informazioni aggiuntive riguardanti ogni categoria di contenuti e comportamenti vietati dalle nostre [Linee Guida per la Community](#).

Oltre ai nostri Termini, forniamo informazioni e risorse relative alla sicurezza nel nostro [Centro per la sicurezza](#), inclusa una guida su [come segnalare potenziali violazioni](#) dei nostri Termini o altri problemi di sicurezza riscontrati sui nostri servizi.

Tutti i documenti sopra menzionati sono allegati alla presente relazione in inglese e sono disponibili sul nostro sito web nelle dodici lingue più parlate da persone con una conoscenza limitata dell'inglese nella città di New York, e in cui è disponibile Snapchat.

Politiche e pratiche di moderazione dei contenuti (NY Gen. Bus. L. § 1102)

I nostri Termini vietano le categorie di contenuti a cui si fa riferimento nel **NY Gen. Bus. L. § 1102(c)**, ossia:

Categoria di contenuti	Categoria di contenuti corrispondente vietata dalle nostre Linee Guida per la Community	Definizioni e politiche pertinenti, come previsto nel nostro Glossario per la Relazione sulla trasparenza e nella nostra serie esplicativa sulle Linee Guida per la Community
Incitamento all'odio o razzismo	Incitamento all'odio (che rientra nella categoria contenuti d'odio, terrorismo ed estremismo violento)	Si definisce incitamento all'odio un contenuto volto a umiliare, diffamare o favorire la discriminazione o la violenza sulla base di razza, colore, ceto sociale, etnia, nazione di provenienza, religione, orientamento sessuale, identità di genere, disabilità o stato di veterano, stato d'immigrato, stato socio-economico, età, peso o stato di gravidanza. Queste norme vietano, ad esempio, l'uso di insulti razziali, etnici, misogini od omofobi. Sono anche vietati i meme che ridicolizzano o invitano a discriminare un gruppo protetto, nonché il deadnaming o il misgendering. L'incitamento all'odio comprende anche il sostegno agli autori di tragedie umane (come il genocidio, l'apartheid o la schiavitù) o il disprezzo delle vittime. Altri contenuti vietati che incitano all'odio sono: l'uso di simboli d'odio, ovvero una simbologia volta a rappresentare l'atto di odiare o discriminare attivamente gli altri. Per ulteriori informazioni, consulta la nostra Guida sui contenuti che incitano all'odio, sul terrorismo e sull'estremismo violento .
Estremismo o radicalizzazione	Terrorismo ed estremismo violento (che rientrano nelle categorie contenuti d'odio, terrorismo ed estremismo violento).	Contenuti che promuovono il terrorismo o altri atti criminali violenti commessi da individui o gruppi per promuovere i propri obiettivi ideologici. Queste regole vietano anche qualsiasi contenuto che promuova o sostenga organizzazioni terroristiche straniere o gruppi che fomentano l'odio estremo, come indicato da esperti credibili di terze parti. Inoltre, è vietato avviare attività di reclutamento per tali organizzazioni o per attività estremistiche violente. Per ulteriori informazioni, consulta la nostra Guida sui contenuti che incitano all'odio, sul terrorismo e sull'estremismo violento .

Disinformazione o informazioni false	Informazioni false (che rientrano tra le pratiche mendaci o ingannevoli).	Include contenuti falsi o fuorvianti che causano danni o hanno natura diffamatoria, come la negazione dell'esistenza di eventi tragici, affermazioni mediche infondate, la compromissione dell'integrità delle procedure civiche o la manipolazione di contenuti per scopi ingannevoli o fuorvianti (anche attraverso l'uso dell'IA generativa o dell'editing ingannevole). Per maggiori informazioni, consulta la nostra Guida sulle pratiche mendaci o ingannevoli .
Molestie	(1) Molestie e atti di bullismo e (2) Molestie sessuali (che rientrano nella sezione contenuti sessuali) (collettivamente, "molestie" nella tabella seguente).	Si riferisce a qualsiasi comportamento indesiderato che potrebbe causare stress emotivo a una persona comune, come abusi verbali, minacce o qualsiasi comportamento volto a far vergognare, mettere in imbarazzo o umiliare una persona. Per ulteriori informazioni, consulta la nostra Guida su molestie e atti di bullismo . Separatamente, le nostre Linee Guida per la Community relative ai contenuti sessuali vietano qualsiasi forma di molestia sessuale. Ciò può includere avance indesiderate, la condivisione di contenuti espliciti e non richiesti o l'invio di richieste oscene o inviti di natura sessuale ad altri utenti. Per ulteriori informazioni, consulta la nostra Guida sui contenuti sessuali .
Interferenze politiche straniere	Informazioni false (che rientrano tra le pratiche mendaci o ingannevoli e includono, a titolo esemplificativo ma non esaustivo, fonti straniere o politiche).	Per la nostra definizione di informazioni false, vedi sopra. Questa categoria include anche il furto d'identità, che si verifica quando un account finge di essere associato a un'altra persona o brand. Per maggiori informazioni, consulta la nostra Guida sulle pratiche mendaci o ingannevoli .

Le nostre [guide esplicative sulla moderazione, sull'applicazione delle norme e sui ricorsi](#) e [quelle sui danni gravi](#) forniscono informazioni dettagliate, tra le altre cose, su:

- come moderiamo i contenuti attraverso strumenti automatizzati e la revisione umana,
- come rispondiamo alle segnalazioni degli utenti su presunte violazioni delle nostre Linee Guida per la Community e
- come interveniamo nei confronti di singoli contenuti e utenti che violano le nostre Linee Guida per la Community.

Informazioni sulla violazione dei nostri Termini (1° luglio - 30 settembre 2025) (NY Gen. Bus. L. § 1102)

Di seguito forniamo informazioni dettagliate sulle violazioni delle nostre Linee Guida per la Community che ci sono state segnalate in-app o che sono state rilevate automaticamente dai nostri sistemi nel periodo dal 1° luglio al 30 settembre 2025, in conformità con la NY Gen. Bus. L. § 1102. I dati forniti sono globali. I seguenti dati escludono le segnalazioni effettuate al di fuori dell'app Snapchat (ad esempio, tramite il sito di assistenza e l'indirizzo e-mail), che rappresentano meno dell'1% del volume totale delle segnalazioni.

Salvo ove diversamente specificato, i termini utilizzati in questa sezione sono definiti in conformità al nostro [Glossario per la trasparenza](#).

Categoria di violazione	Modalità di segnalazione	Totale contenuti o account segnalati ⁽¹⁾	Contenuti moderati ⁽²⁾ da revisori umani	Contenuti moderati da strumenti automatizzati	Account unici moderati ⁽³⁾ da Human Reviewers	Account unici moderati da strumenti automatizzati	Ricorsi contro blocchi dell'account valutati da revisori umani	Ricorsi contro blocchi dell'account valutati da strumenti automatizzati	Account ripristinati a seguito di ricorso ⁽⁴⁾ (inizialmente bloccati da revisori umani)	Account ripristinati a seguito di ricorso (inizialmente bloccati da strumenti automatizzati)	Violative View Rate (VVR) ⁽⁵⁾ per i contenuti moderati da revisori umani	VVR per i contenuti moderati da strumenti automatizzati	Tasso di visualizzazione di violazioni uniche ⁽⁶⁾ per i contenuti moderati da revisori umani	VVR unico per i contenuti moderati da strumenti automatizzati
Incitamento all'odio	Segnalazione umana	482.240	144.389	20.679	123.835	18.343	94	0	5	0	0,000205%	0,000004%	0,41%	0,01%
	Rilevamento proattivo	3.073	2.441	32	2.029	29	7	1.164	1	15	0,000002%	0,000000%	0,00%	0,00%
Terrorismo ed estremismo violento	Segnalazione umana	199.245	559	54	468	54	199	0	1	0	0,000037%	0,000000%	0,07%	0,00%
	Rilevamento proattivo	12.240	5.901	0	3.754	0	11	38	1	6	0,000000%	0,000000%	0,00%	0,00%
Informazioni false	Segnalazione umana	386.210	145	593	113	593	1	0	0	0	0,000003%	0,000000%	0,01%	0,00%
	Rilevamento proattivo	262	10	0	10	0	1	0	0	0	0,000000%	0,000000%	0,00%	0,00%
Molestie	Segnalazione umana	3.184.734	756.736	604.798	624.142	491.972	700	0	17	0	0,001135%	0,000168%	1,93%	0,35%
	Rilevamento proattivo	12.293	8.972	1.583	7.670	1.428	17	13.442	1	97	0,000007%	0,000000%	0,02%	0,00%

- (1) Numero totale di contenuti o account che sono stati segnalati per potenziali violazioni delle nostre Linee Guida per la Community, compresi quelli segnalati dagli utenti e quelli rilevati attraverso i nostri processi di rilevamento proattivo. Per disaggregare questi dati in categorie di contenuti illeciti, abbiamo utilizzato il motivo ultimo per cui è stata intrapresa un'azione di contrasto. Laddove il contenuto o account sia stato segnalato ma non sia stata intrapresa alcuna azione restrittiva, attribuiamo i dati alla categoria di sospetta violazione per la quale il contenuto o account è stato segnalato.
- (2) Il numero di contenuti (ad esempio, Snap, Storie) che sono stati moderati su Snapchat. Per "moderazione" si intende un'azione intrapresa nei confronti di un contenuto o di un account (ad esempio, eliminazione, avviso, blocco).
- (3) Il numero di account univoci che sono stati moderati su Snapchat. Ad esempio, se un singolo account è stato sanzionato più volte per vari motivi (ad esempio, un utente ha ricevuto un avvertimento per aver pubblicato informazioni false e poi è stato bloccato per aver molestato un altro utente), solo un account sarà calcolato in questa metrica. Come sopra, "moderazione" si riferisce a un'azione intrapresa nei confronti di un contenuto o di un account (ad esempio, eliminazione, avviso, blocco).
- (4) Ripristiniamo un account solo se i nostri moderatori stabiliscono che è stato bloccato ingiustamente.
- (5) Il Violative View Rate indica la percentuale di visualizzazioni di Storie e Snap che contenevano contenuti non conformi, rispetto a tutte le visualizzazioni di Storie e Snap su Snapchat. (uno Snap è una foto o un video scattato con la fotocamera di Snapchat. Maggiori informazioni sono disponibili [qui](#)). Ad esempio, se il nostro VVR è pari allo 0,03%, significa che ogni 10.000 visualizzazioni di Snap e Storie su Snapchat, 3 contenevano contenuti che violavano le nostre politiche. Questa metrica ci consente di capire quale percentuale di visualizzazioni su Snapchat proviene da contenuti che violano le nostre Linee Guida per la Community (che sono stati segnalati o fatti rispettare in modo proattivo).
- (6) Il Violative View Rate unico è la percentuale di utenti unici che hanno visualizzato le Storie e/o gli Snap non conformi, rispetto agli utenti unici attivi durante il periodo di segnalazione (ad esempio, il terzo trimestre del 2025). Ad esempio, se il nostro Violative View Rate unico è pari allo 0,03%, significa che, per ogni 10.000 utenti attivi durante il periodo di riferimento su Snapchat, 3 utenti hanno visualizzato Storie e/o Snap che hanno violato le nostre politiche. Questa metrica ci consente di capire quale percentuale di utenti su Snapchat si imbatte in Storie e/o Snap che violano le nostre Linee Guida per la Community (che sono stati segnalati o moderati in modo proattivo).

Informazioni aggiuntive

Per ulteriori informazioni sull'approccio di Snap alla sicurezza, alla privacy e alla trasparenza, visita il nostro [Hub sulla privacy, la sicurezza e le politiche](#) e la nostra [pagina sulla relazione per la trasparenza](#).