Content Moderation Policy

CreamyPeach (creamypeach.com) is operated by **PEACHCREAM S.R.L.**, a company registered in Romania (Trade Register no. J2026041854006, CUI 55048995), with its registered office at Șoseaua Iancului nr. 53, bl. 102B, sc. C, et. 7, ap. 111, Sector 2, Bucharest, Romania ("**CreamyPeach**", "**we**", "**us**", "**our**").

Version: 1.1 | **Effective date:** 30 October 2026 | **Contact:** support@creamypeach.com

This Policy describes how we keep prohibited content off CreamyPeach: what is reviewed, by whom, with which tools, before and after publication, and what happens when something is found. It forms part of our Platform Policies and should be read together with our Acceptable Use Policy (which defines prohibited content), our Identity and Age Verification Policy, our Complaints, Takedown and Appeals Policy and our Anti-Human Trafficking and Exploitation Policy.

1. Principles

1. **Nothing is published without review.** Every image, video, audio file, post, profile element and pay-per-view item is reviewed by a trained person **before** it becomes visible to anyone other than the uploader. There is no "publish first, review later" on CreamyPeach. Private text messages between fans and creators, and messages from AI characters, are not "published": they are screened automatically in real time before delivery, with human review of every message that is flagged and of a random sample of all messages.
2. **Humans decide.** Automated tools screen, flag, block and prioritise. Decisions to publish, reject or remove content, and every suspension or termination of an account, are made or confirmed by trained people.
3. **AI content is held to the same standard.** Content generated by artificial intelligence — whether by CreamyPeach's own AI characters or by creators — is subject to the same review and the same prohibitions as content depicting real people, and to additional guardrails designed for synthetic content.
4. **Verification before content.** No content depicting a person is published unless that person has been verified as an adult and has consented. See our Identity and Age Verification Policy.
5. **Zero tolerance.** Content involving minors, non-consensual content, synthetic depictions of real people other than a creator's own consented persona, bestiality and human trafficking is removed immediately, the account is permanently terminated, and the matter is reported to the authorities where a crime is indicated.
6. **Transparency.** Users whose content is rejected or removed are told why and how to appeal.

2. What we moderate

All content on CreamyPeach, specifically:

- creator profiles, usernames, display names, biographies, avatars and banners;
- feed posts (images, videos, audio, text and captions);
- pay-per-view items and their previews and descriptions;
- direct messages between fans and creators, including messages generated by AI characters;
- prompts and inputs that users send to AI characters or to any generation tool;
- all AI-generated images, videos, audio and text, before they enter our content library; and
- tags, hashtags, links and any other metadata.

3. Moderation layers

Layer 1 – Prevention at the source

- **Verification gate.** Only verified adults can create accounts; only fully identity-verified creators can upload; only verified, consenting co-performers can appear in content.
- **AI guardrails.** Our AI characters and generation pipeline are governed by an internal AI Content Guardrails Policy that blocks prohibited categories at the point of generation: prohibited themes are excluded from character design, from prompts, from the models used, and from outputs. AI characters are fictional, created and presented as adults, and are not based on any real person, except that a verified real creator may operate an AI persona of their own likeness under the Acceptable Use Policy. All AI content is labelled as artificially generated, and AI characters identify themselves as artificial in every conversation.
- **Prohibited-term filter.** Usernames, profile text, captions, tags, messages and prompts are screened against a multilingual list of prohibited terms and patterns (including terms referencing minors, family relationships in a sexual context, non-consent, animals, bodily waste, escort or in-person services, and known evasion spellings). Matches are blocked or held for human review, depending on severity.
- **Creator education.** Every creator must read and accept the Acceptable Use Policy and Creator Agreement before uploading and receives clear in-product guidance on what is prohibited.

Layer 2 – Automated screening of every upload and generation

Every file is passed through automated screening before it reaches a human reviewer:

- **Known-abuse matching.** Image and video hashes are compared against databases of known child sexual abuse material and other prohibited material operated by recognised child-protection organisations. A match results in immediate blocking, preservation, account termination and reporting.
- **Machine-learning classification.** Images, video frames and audio are classified for: indicators of a minor (apparent age, body characteristics, settings, props, clothing); indicators of non-consent, violence, injury, intoxication or unconsciousness; animals; bodily waste; weapons; family-themed scenarios; and other prohibited

categories. Text (captions, messages, prompts, AI outputs) is classified for the same categories.

- **Person count and release matching.** For content depicting real people, the number of persons detected in the content is compared with the verified persons registered for that content (the creator and each co-performer who has signed a release). Content with more persons than registered releases is held. CreamyPeach does not hold biometric templates and does not perform automated biometric face matching; identity is established through the verification provider at onboarding and through the signed release linked to each item of content.
- **Synthetic-media checks.** AI-generated content is checked for the presence of our AI label and provenance metadata and is screened for resemblance to real public figures through a third-party likeness-screening service, and to other verified creators through our own verification records. CreamyPeach does not build its own database of real people's faces.
- **Risk scoring.** Each item receives a risk score. Items scoring above the "block" threshold are rejected automatically and escalated; all other items proceed to human review with their risk signals attached. Automated screening never publishes anything on its own.

Layer 3 – Human pre-publication review

- Every item is reviewed by a trained member of our Trust & Safety team before publication, with the automated signals displayed alongside it.
- Reviewers apply a written Pre-Publication Review Checklist derived from the Acceptable Use Policy, card-network standards and our AI Content Guardrails Policy, and they must confirm, for content depicting real people, that each person has a matching verification record and a signed Co-Performer Release and Consent Form.
- Reviewers reject, approve, or escalate. Any item where a reviewer has doubt about the age of a person depicted, or about consent, is rejected and escalated; it is never approved "on balance".
- Rejected items are returned to the creator with the reason and the right of appeal.

Layer 4 – Post-publication monitoring

- **Continuous rescanning.** Published content is periodically rescanned as classifiers and abuse databases are updated.
- **Message monitoring.** Direct messages and AI-character conversations are screened in real time by the prohibited-term filter and text classifiers. High-risk messages are blocked before delivery and escalated; medium-risk messages are delivered and queued for review.
- **Behavioural signals.** We monitor account-level patterns (device, IP and payout links between accounts, sudden changes in content type, repeated near-miss rejections, fan complaints, chargebacks) and review accounts that show indicators of risk.
- **Reports.** Every user can report any content or account using the Report button or by email. Reports are triaged and handled under our Complaints, Takedown and

Appeals Policy, with urgent categories actioned immediately, content removed within 48 hours at the request of a person depicted, and all reports resolved within seven business days.

- **Trusted sources.** Reports from law enforcement, regulators, child-protection hotlines and other trusted organisations are prioritised.

Layer 5 – Escalation and specialist review

- Items flagged as possible child sexual abuse material, non-consensual content, trafficking or coercion are escalated to our Trust & Safety lead, who confirms the removal, the account action, the preservation of evidence and the reports to authorities and payment partners.
- Novel or borderline cases are decided by the Trust & Safety lead and recorded, with the reasoning, in our internal moderation guidance so that decisions are consistent.

4. Enforcement actions

Finding	Content action	Account action	Reporting
Child sexual abuse material or sexualised depiction of a minor (real or AI)	Immediate removal and preservation	Permanent termination; identity, device and payment method blocked	Romanian Police, NCMEC CyberTipline and other competent authorities; payment partners
Non-consensual intimate imagery, deepfake or likeness without consent	Immediate removal and preservation	Permanent termination	Law enforcement where a crime is indicated; payment partners
Bestiality	Immediate removal	Permanent termination	Law enforcement where a crime is indicated; payment partners
Human trafficking, coercion, prostitution or escort activity	Immediate removal and preservation	Permanent termination; payouts withheld	Romanian Police, DIICOT, ANITP and other competent authorities; payment partners
Incest themes, violence, bodily waste, hate, illegal activity or other prohibited content	Removal	First confirmed breach: warning and strike. Second: 30-day suspension. Third: permanent termination. Severe cases: immediate termination	As required by law
Unverified person	Content held and	Warning;	As required by law

Finding	Content action	Account action	Reporting
appearing in content	not published	termination for repeat or deliberate breaches	
Attempts to circumvent AI guardrails or prohibited requests to AI characters	Prompt blocked; output withheld	Warning and strike for a first attempt; 30-day suspension for a second; termination for a third. Immediate termination for any request involving minors, animals, or in-person sexual services or trafficking	Requests involving minors may be reported to the authorities
Other breaches of the Acceptable Use Policy	Removal or edit required	Graduated: warning, strike, suspension, termination	—

Earnings connected to a confirmed serious breach are withheld and, to the extent permitted by law, forfeited. Automated systems may block content or suspend an account instantly; every termination is confirmed by a trained member of our team within 24 hours. All enforcement decisions are notified to the user with reasons and can be appealed under our Complaints, Takedown and Appeals Policy, except where that Policy excludes appeal.

5. Our Trust & Safety team

- Moderation is performed by trained members of our team under the direction of the **Trust & Safety lead**, who reports to the Chief Executive Officer of PEACHCREAM S.R.L.
- All reviewers complete training before reviewing any content, covering the Acceptable Use Policy, card-network standards for adult content, recognition of indicators of minors, non-consent, coercion and trafficking, the handling and reporting of child sexual abuse material, and our escalation and record-keeping procedures. Refresher training takes place at least annually and whenever policies change.
- From launch, reviewer decisions are sampled and audited weekly by the Trust & Safety lead for accuracy and consistency.
- We provide reviewers with wellbeing support appropriate to the nature of the work.
- We maintain moderation capacity so that the pre-publication queue is cleared within one business day under normal conditions.

6. Testing and assurance

- We test our guardrails and moderation systems against each prohibited category **at least quarterly**, using a documented methodology, and we record the results in a written attestation signed by our Chief Executive Officer. Any failure identified is remediated before the attestation is signed.
- We engage independent third parties to assess our AI content guardrails where required by our payment partners.
- We maintain an incident-response process under which confirmed incidents are escalated internally within hours and notified to our payment partners within 24–48 hours of detection.

7. Records and transparency

We keep records of every moderation decision, including the content concerned, the automated signals, the reviewer, the decision, the reasons, the time taken and any appeal. Records are retained for as long as necessary to meet our legal obligations, to support investigations by competent authorities and to satisfy the requirements of our payment partners. From launch, we publish periodic transparency information about the volume of content reviewed, rejected and removed, and about reports and appeals received, and we make detailed reports available to our payment partners on request.

8. Cooperation with authorities

We cooperate fully with law enforcement and regulators. We report child sexual abuse material and suspected trafficking to the competent authorities as described above, we preserve evidence, and we respond promptly to lawful requests for information. Where we become aware of information giving rise to a suspicion that a criminal offence involving a threat to the life or safety of a person has taken place, is taking place or is likely to take place, we promptly inform the law enforcement or judicial authorities of the Member State or State concerned.

9. Changes to this Policy

We may update this Policy from time to time, in particular to improve our tools, to reflect changes in law or card-network rules, or to meet the requirements of our payment partners. The current version is always available at creamypeach.com. Material changes will be notified to registered users by email or in-platform notice at least 14 days before they take effect, unless a change is required by law or to address an urgent safety risk, in which case it takes effect immediately.

10. Contact

PEACHCREAM S.R.L.
Șoseaua Iancului nr. 53, bl. 102B, sc. C, et. 7, ap. 111, Sector 2, Bucharest, Romania
Email: support@creamypeach.com (reports) · legal@creamypeach.com (legal notices) · dsa@creamypeach.com (Digital Services Act authorities)