

UGC Moderation Pipeline — Build Blueprint

Fora Soft Learn · AI for Video Engineering · Capstone 10.6

1 · The pipeline: two rules, five stages (a funnel)

- **Rule 1** — Cheapest check first — each stage decides or escalates; expensive stages never see the bulk.
- **Rule 2** — Every decision logged with a reason; illegal content reported + preserved, not deleted.
- **Stages** — Ingest → Stage0 hash → Stage1 classifiers → Stage2 VLM → Stage3 human, over a governance floor.
- **Key move** — Moderate sampled keyframes + audio transcript, NOT whole videos by the minute.

2 · Build vs buy (2026): rent the models, partner on hashes, build your policy

- **BUY / HOST** — classifiers (Rekognition/Hive/Azure/NudeNet/CLIP) + VLM (Gemini/Claude/GPT/LLaVA/Qwen-VL)
- **PARTNER** — known-bad hash systems (PhotoDNA, NCMEC/IWF, PDQ+TMK, Lantern) — never build/hold yourself
- **BUILD** — routing + thresholds + decision store + audit + reporting + humane review tooling = product
- **ABSTRACT** — classifier and VLM behind one interface each, so a vendor swap is a config change.

3 · The 2026 tool field (verify prices before you ship)

- **CLASSIFIERS** — Azure ~\$1.50/1k images (~\$0.0015/img) | Rekognition ~\$0.10/min video | Google | Hive | open
- **VLM** — frontier Gemini/Claude/GPT for the hard cases | open LLaVA/Qwen-VL for routine context
- **HASH** — PhotoDNA + NCMEC/IWF sets + Meta PDQ/TMK + Tech Coalition Lantern/VHIP (784k+ videos hashed)
- **NOTE** — hash catches the KNOWN only; AI-generated CSAM has no prior hash → classifiers + VLM matter.

4 · The cost model (per day at 100,000 videos)

- **NAIVE** — 200,000 video-min/day × \$0.10/min = \$20,000/day ~ \$7.3M/year (and too slow).
- **FUNNEL** — hash \$50 + classifiers \$1,800 + VLM(~7%) \$140 + humans(~1%) \$250 = ~\$2,240/day ~ \$0.8M/year.
- **RESULT** — ~9× cheaper AND more accurate — the costly stages only see what truly needs them.
- **LEVER** — tune Stage 1 thresholds (the share it clears), not the model, to win the bill.

5 · Build order: each milestone protects users

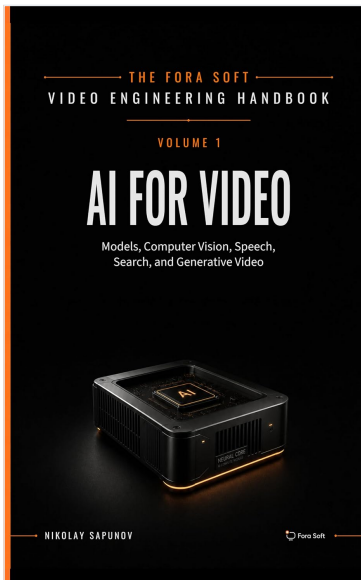
1. Hash matching + mandatory reporting (Stage 0 + legal floor) — cheapest and legally required, first.
2. Cheap classifier pass, conservative thresholds (Stage 1) — escalate generously, tighten with evidence.
3. Human review console + queue (Stage 3) BEFORE the VLM — with wellbeing features built in from day one.
4. VLM context stage (Stage 2) to shrink the human queue.
5. Scale + measurement + adversarial loop.

6 · Governance checklist (engineering context, not legal advice)

- ☐ Every chunk carries source + timecode from ingestion, so every decision can be logged with its reason.
- ☐ On a Stage 0 match: BLOCK + REPORT (NCMEC CyberTipline) + PRESERVE — never a quiet delete (18 U.S.C. 2258A).
- ☐ Thresholds set PER HARM CLASS — gravest harms deterministic (hash + human), not a tunable score.
- ☐ DSA statement-of-reasons + transparency; minors protected (Art. 28); EU AI Act Art. 50 marks AI media (Aug 2026).
- ☐ Reviewer wellbeing as a requirement: blurred previews, hard exposure caps per shift, rotation, living wage.
- ☐ Held-out known-answer eval set run continuously; novel confirmed harms fed back to the hash authorities.

Companion to: 'Capstone — UGC Moderation Pipeline at 100,000 videos/day'

www.forasoft.com/learn/ai-for-video-engineering/articles-ai/ugc-moderation-100k-day-scale



GO DEEPER · VOLUME 6 OF 7

AI for Video: A Complete Guide to Intelligence in Video Products

From Models and Inference to Computer Vision, Speech, Search, and Generative Video

By Nikolay Sapunov, CEO at Fora Soft · 1,444 pages · Kindle & paperback on Amazon

[Read it on Amazon →](#)
amazon.com/dp/B0HH6TKPZ1

This blueprint lays out the moderation funnel, the build-versus-buy split, the cost model and the governance floor. Volume 6 covers the capabilities inside each stage: computer vision for detection and classification, speech recognition for audio transcripts, search over a video archive, and deployment topologies that keep cost per video down.

WHAT YOU WILL FIND INSIDE

- Computer vision: detection and classification
- Speech recognition for transcripts
- Search and RAG over a video archive
- Deployment topologies and AI feature lifecycle

THE FULL HANDBOOK · 7 VOLUMES · 9,000+ PAGES · ALL ON AMAZON

Vol. 1 · Video Encoding — pixels and color to H.264, HEVC, AV1, quality metrics and AI

Vol. 2 · Video Streaming I — network transport, contribution and ingest, HLS, MPEG-DASH, CMAF

Vol. 3 · Video Streaming II — ABR, CDNs, players, WebRTC, DRM and the economics of delivery

Vol. 4 · Audio for Video I — PCM, codecs, loudness, accessibility, real-time voice processing

Vol. 5 · Audio for Video II — jitter buffers, lip sync, Atmos, quality metrics, neural codecs

Vol. 6 · AI for Video I — models and inference, computer vision, speech, search, generative video

Vol. 7 · AI for Video II — real-time pipelines, agents, serving, cost, compliance, verticals

Need to moderate user video at scale without a seven-figure bill?

We will design the hash, classifier, VLM and human-review funnel for your volume, set thresholds per harm class, and model the cost per day. 250+ video products shipped since 2005.

[Book a 30-min call →](#)
[WhatsApp →](#)
[Email us →](#)