

1 - The three gates between a model and a user

QUALITY (judge: pass / retry / human) -> COST (cap per request, per user, per system)
-> DISCLOSURE (C2PA manifest + watermark + log + label). Every gate fails closed.

2 - Quality gate (cheapest check first)

- Technical validation: file decodes, right length / resolution / frame rate, brightness sane.
- Cheap perceptual checks: face + finger detection, no-reference quality score.
- VLM judge on sampled frames: prompt match, intact faces, readable correct text, stable motion.
- Hard retry cap (e.g. 3); on cap, fail loudly or escalate to a human - never loop forever.

3 - Cost gate (three caps)

- Per-request cap: max clip length, max resolution, max retries.
- Per-user cap (credits): credits = monthly margin target / cost per delivered clip.
- Per-system cap: total-spend circuit breaker with alert + automatic throttle.

Worked example: \$8.00 / mo budget / \$0.50 per clip = 16 credits per user / month

4 - Disclosure gate (EU AI Act Article 50, in force 2 Aug 2026)

- Mark every AI-generated output machine-readably (Art. 50(2) - providers).
- Disclose any deepfake: AI-made/alterd video of real people, places, or events (Art. 50(4) - deployers).
- Embed C2PA Content Credentials: signed manifest (generator, time, signer) + SHA-256 hard binding.
- Add an invisible watermark (e.g. SynthID) that survives crop, re-encode, and compression.
- Keep a server-side log: clip, model, user, prompt, consent, disclosure - your evidence of record.
- Show a clear 'AI-generated' label at first exposure; capture consent BEFORE generation.

Penalty for breaching Article 50: up to EUR 15,000,000 or 3% of worldwide annual turnover.