

1 - The five-stage pipeline (every tool runs this)

1 Transcribe (ASR, word timestamps) -> 2 Detect moments (LLM) -> 3 Reframe (CV face track)
-> 4 Captions (word-level, burned-in) -> 5 Render (cut, composite, encode MP4).

2 - The five tools mapped to the pipeline

Tool	Owns	2026 entry	Best for
Opus Clip	Moments + reframe	\$15/mo	Automated clipping, has API
Descript	Edit as transcript	\$16/mo	Full episodes, courses, podcasts
Submagic	Captions + B-roll	~\$12/mo	Animated captions (30-min cap)
Captions (Mirage)	Restyle + presenter	\$9.99/mo	AI avatars, one-tap restyle
DaVinci Resolve 20	Assistive AI in editor	Free	Editors keeping manual control

Prices are 2026 list ranges and change often - re-check at integration time. Tools are combined, not chosen exclusively.

3 - Build the pipeline or buy the API?

BUY a vendor API when volume is low: no infra, free tier + flat price beat your engineering time.
BUILD the pipeline when volume is high (1000s of clips/mo) OR you need styles the vendor lacks
or cannot send user video out. Hidden build cost: an LLM runs on EVERY video (~18k tokens / 30 min).

4 - Two numbers to remember

Word-timestamp accuracy: ~30-40 ms on clean audio, 80-100 ms on noisy. Captions cap on worst audio.
Show 4-6 caption words at a time max. ASS karaoke tags + FFmpeg burn-in; SRT cannot animate words.

5 - EU AI Act disclosure checklist (in force 2 Aug 2026)

- Mark every AI-generated output in a machine-readable format detectable as AI-made.
- Disclose any video that is a deepfake - AI-generated or AI-manipulated (avatars, regenerated voice).
- Embed C2PA Content Credentials: a signed manifest of which AI made it, when, and who signed.
- Add an imperceptible watermark that survives compression, cropping, and screenshots.
- Keep a consent + audit trail; treat disclosure as a pipeline station, not an afterthought.