

1 - Two tracks (sort every tool first)

Track A	AI inside the product - features users touch. On the product + compliance budget.
Track B	AI for building the product - tools engineers use. On the productivity budget.
The test	Does a paying customer experience the output? Yes = Track A. No = Track B.

2 - The five-stage lifecycle map (where a Track A tool attaches)

- 1 Plan & ingest script / storyboard draft; archive tagging & search.
- 2 Produce / capture LIVE, on a clock: captions, translation, noise & blur, moderation.
- 3 Post-produce no clock, cheapest: clipping, dubbing, subtitles, voice-over, b-roll.
- 4 Deliver per-title encoding, archive upscaling, personalization.
- 5 Operate & govern . eval, rigs, cost control, provenance & disclosure.

3 - Four ways to source it + the buy-vs-self-host crossover

Buy a hosted API . fastest, per-minute cost, data leaves your walls; low / spiky volume.
Fine-tune a small model beats a big one on a narrow domain; needs a labeled set.
Self-host fixed cost, near-free per minute; high steady volume, strict privacy.
Build custom only when the capability IS the product and nothing off-the-shelf fits.
Crossover = fixed self-host cost / API per-minute price -> below it rent, above it own.

4 - Stable spine vs swappable model + the four cost lines

SPINE (years): lifecycle stages, delivery standards, browser media, your own interfaces.
SWAP (weeks): the specific model, the specific vendor, prices - wrap each in one interface.
Cost lines: model fee (smallest) + integration + evaluation & quality + governance.

5 - Governance + scoping gate - before you build (context, not legal advice)

- Sort every requested tool into Track A or Track B at scoping; budget them on separate ledgers.
- Place each Track A capability on a lifecycle stage; keep 'later' work off the live (clock) path.
- Pick a sourcing mode by real volume and privacy; compute the buy-vs-self-host crossover first.
- Wrap every model call behind your own interface so the vendor's name lives in exactly one file.
- Budget all four cost lines - the model fee is usually the smallest in year one.
- AI that generates or substantially alters video: mark output (EU AI Act Art 50, from 2 Aug 2026) via C2PA; sign at generation time.