

1 - Pick the lightest mode that does the job

SUBTITLES (original voice + on-screen text; cheapest, no cloned voice, lowest risk)
VOICE-OVER (narrator over ducked audio; docs/news) FULL DUB (cloned + lip-sync; = deepfake)

2 - One pipeline, three outputs (front half is shared)

TRANSCRIBE (ASR + word timestamps + speaker labels) -> TRANSLATE (length-controlled) -> branch:
Subtitles: typeset to CPS | Voice-over: TTS -> timing -> mux | Dub: clone -> timing -> lip-sync -> replace

3 - Auto-subtitle rules (split every cue until all pass)

- Reading speed: ~17 CPS adult, ~13 CPS children. A 3.0 s cue x 17 CPS = 51 characters maximum.
- Max 42 characters per line; max 2 lines per cue; duration 5/6 s minimum to 7 s maximum.
- Break lines at clause boundaries, not mid-phrase; never ship the raw ASR transcript as subtitles.
- Deliver WebVTT for web/HLS/DASH; deliver IMSC1 (W3C Rec.) for broadcast/OTT in fMP4/CMAF.

4 - Quality gate (a number at every stage; fail to a human)

- Transcription: word error rate (WER) ceiling = (subs + deletions + insertions) / words.
- Translation: COMET floor (neural metric) for the real gate; BLEU only for quick regression.
- Voice: mean opinion score (MOS, ITU-T P.800) target near 4.0; tighter for lip-synced dialogue.
- Timing: every dub clip within +/-10-15% time-stretch; bilingual human spot-check on a sample.

5 - Disclosure (synthetic voices) + accessibility (captions)

- Cloned voice of a real person = audio deepfake: capture consent BEFORE cloning and store it.
- EU AI Act Art. 50 (in force 2 Aug 2026): mark synthetic audio machine-readably; disclose deepfakes.
- Show a clear 'AI-generated voice' label at first exposure; keep a server-side disclosure log.
- FCC caption quality: accuracy, synchronicity, completeness, placement; CVAA for TV re-runs online.
- Meet WCAG captions for prerecorded video; the CPS and placement rules above are the legal core.

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