

1 - Two jobs under one name (decide this first)

LIP-SYNC: start from real video, redraw ONLY the mouth. Cheaper, faster, safer-looking.
 AVATAR GEN: start from ONE photo, invent the whole head. More powerful, heavier, riskier.

2 - The five open-source models

Model	Job	Real-time?	Best for
Wav2Lip	Lip-sync	Yes	Any face, cheap coverage
MuseTalk	Lip-sync	Yes	Real-time mouth on video (~30 fps)
LatentSync	Lip-sync	No	Highest-quality lip-sync (diffusion)
SadTalker	Avatar gen	No	Avatar from one photo (3D scaffold)
EMO / OmniHuman	Avatar gen	No	Expressive offline avatars (diffusion)

3 - The three commercial platforms

Platform	Built for	2026 entry
Tavus	Live AI video agents (CVI, sub-600 ms)	Free, then ~\$59/mo
HeyGen	Scripted multi-language presenters + API	Free, then ~\$29/mo
Synthesia	Governed enterprise training libraries	~\$29/mo (~\$22 annual)

Prices are 2026 list ranges and change often - re-check at integration time.

4 - The real-time line (why diffusion can't stream live)

30 fps -> 1 s / 30 = 33 ms per frame. Every frame must finish inside 33 ms or it stutters.
 A 25-step diffusion model: 33 / 25 = 1.3 ms per step. No stock diffusion runs a step that fast.
 => Real-time = single-step methods only. EMO-class output is rendered OFFLINE, then played back.

5 - Consent & disclosure checklist (before you ship)

- Store WHO consented, to exactly WHAT use, WHEN - with an expiry and a revocation path.
- Tie every generated asset back to the consent record that authorized it (audit trail).
- Get a recorded consent statement before cloning any voice or face - even if the vendor doesn't force it.
- Add a visible 'AI-generated' disclosure - EU AI Act Article 50 binds deepfake disclosure on 2 Aug 2026.
- Track the NO FAKES Act (US, proposed 2025): a federal right over a person's digital replica is coming.