60-second takeaway
For our short social-video dubbing workflow, LatentSync 1.6 is still the local model to beat. KeySync is the closest direct A/B alternative because it solves the same existing-video plus replacement-audio task. MuseTalk remains useful when speed matters, but our current evidence does not put it above LatentSync 1.6. InfiniteTalk and LTX LipDub are the interesting next wave, but they are heavier talking-video systems rather than simple drop-in mouth replacement tools.

The task matters more than the model name

Lip sync comparisons get messy because teams often mix three different jobs:

- replacing mouth motion in an existing video from new audio
- animating a still portrait from speech
- generating a new talking video where lips, face, head, body, and camera motion are all synthesized together

Those are related, but they are not interchangeable.

For a production social-video pipeline, our first question is narrower:

- can this model take a real source clip and a replacement audio track, then produce a believable dubbed video without wrecking identity, crop, expression, or timing

That is why LatentSync 1.6 remains the baseline in our internal benchmark notes. It is not a universal claim that LatentSync wins every talking-head task. It is a fixture-bound decision for our current dubbing use case.

Current decision

Model	Comparable to LatentSync?	Practical read	Current status
LatentSync 1.6	Baseline	Strong current local baseline for existing-video lip replacement.	Preferred default for our tested use case.

KeySync	Very high	Same core task: align lip movement in an existing video to new audio, with explicit attention to expression leakage and occlusion.	Closest direct A/B fallback, but our initial local verdict preferred LatentSync 1.6.
MuseTalk	High	Real-time, audio-driven lip sync with latent-space inpainting. Useful when speed matters.	Downgraded to speed-oriented fallback only for our current use case.
InfiniteTalk	High, but heavier	Audio-driven video-to-video and image-to-video talking generation with lip, face, head, and body motion.	Future candidate, not a proven replacement for LatentSync in our workflow.
LTX LipDub	Medium-high	LTX 2.3 IC-LoRA for lip dubbing inside the LTX video ecosystem.	Promising watchlist item, not yet a drop-in production default.
VideoReTalking	Medium	Mature multi-stage talking-head editing pipeline.	Useful baseline, likely less attractive than newer models for modern social clips.
Wav2Lip	Medium-low	Classic local baseline with many wrappers and a simple mental model.	Good sanity baseline, but quality and dependency age limit production appeal.
Diff2Lip	Medium	Diffusion-based research direction for audio-conditioned lip synchronization.	Research reference unless the implementation path is easy to operate.
Hallo / Hallo2	Adjacent	Audio-driven portrait animation from an image.	Useful for avatars, not the same as existing-video dubbing.
EchoMimic	Adjacent	Audio-driven portrait animation with landmark conditioning.	Better framed as presenter/avatar animation than video mouth replacement.
TalkVerse	Watchlist	Minute-long audio-driven video-generation research	Needs released-code and checkpoint

	direction based on Wan-style video models.	maturity checks before planning around it.

What our internal benchmark says

Our current repo benchmark policy separates host-level model state from product decisions.

The product decision is simple:

- **Current preferred local baseline:** LatentSync 1.6.
- **Closest direct fallback:** KeySync.
- **Speed-oriented fallback:** MuseTalk.
- **Future candidates:** InfiniteTalk and LTX LipDub.

The important caveat is that this is not a broad public leaderboard. It is a production decision for short social-video dubbing fixtures, where the output has to survive crop, compression, timing, identity checks, and human review.

What breaks in real use

Forum threads and issue trackers make one point clear: lip sync failures are rarely just about whether the lips move on the right syllable. Users also complain about full-frame blur, mouth boxes, artificial teeth, identity drift, jitter, crop damage, sync drift, and setup failures.

That evidence is not a statistical survey. GitHub issues overrepresent developers with broken installs, Reddit threads overrepresent frustrated users, and Product Hunt reviews mix product praise with billing and support complaints. Still, the pattern is useful for production decisions: a model that looks good in a demo can still fail a real social-video workflow.

The recurring failure modes are:

- full-frame quality loss outside the mouth region
- visible mouth boxes, black mouth artifacts, jitter, or unnatural teeth
- identity changes even when lip timing looks accurate
- sync drift on longer clips or videos with cuts

- VRAM, CUDA, PyTorch, xformers, mmcv, Docker, and Windows setup friction
- unclear input requirements around face angle, occlusion, speaker count, and audio duration
- API credit surprises, watermarks, moderation blocks, and support opacity
- confusion between existing-video mouth replacement, portrait animation, and full talking-video generation

That is why our benchmark rubric treats frame preservation, teeth, identity, crop, latency, licensing, and preprocessing burden as first-class criteria. Sync accuracy is necessary, but it is not enough.

KeySync: the closest direct challenger

KeySync is the model I would retest first when asking whether LatentSync 1.6 should be replaced.

The reason is task fit. Its own project describes lip synchronization as aligning lip movement in an existing video with new input audio, and it targets expression leakage and occlusion handling. That maps directly onto the risks we care about in real dubbing:

- mouth timing looks good but cheeks or jaw inherit the wrong expression
- a hand, microphone, subtitle, or product partially occludes the face
- the mouth region syncs while the rest of the face looks inconsistent
- preprocessing turns a vertical social clip into an awkward square crop

Our local result so far did not promote KeySync above LatentSync 1.6. That does not make KeySync irrelevant. It makes it the right challenger for a bounded bakeoff with the same source clips, the same audio, and the same review rubric.

The issue-tracker caveat is that KeySync should not be judged on sync alone. Reported problems include identity drift, artificial teeth, distorted color, and sync errors tied to FPS or sample-rate assumptions. Those are exactly the kinds of failures a fixture bakeoff has to catch before any promotion.

MuseTalk: speed fallback, not current winner

MuseTalk matters because it is designed for real-time, high-quality lip syncing and reports 30 FPS plus performance on an NVIDIA Tesla V100. For product

workflows, that speed signal is not trivial. Fast preview loops can matter more than the final render winner when creators are iterating.

But our current evidence is clear enough for routing:

- MuseTalk is not our preferred quality default.
- It is still worth keeping as a speed-oriented fallback.
- Its code license and model-weight terms should be checked separately before any production commitment.
- Dependency and environment setup should be treated as part of the benchmark, not an afterthought.

The practical conclusion is not "MuseTalk is bad". It is "MuseTalk has to win a speed-constrained fixture to displace LatentSync for that lane."

The public issue pattern matches that routing. MuseTalk users report out-of-sync output, mouth jitter, blurry or black mouth regions, and environment friction around OpenMMLab-style dependencies, PyTorch, mmcv, ffmpeg, Windows, and newer GPU support. That does not erase the speed story, but it means speed must be measured alongside setup burden and visible mouth artifacts.

InfiniteTalk and LTX LipDub: the heavier next wave

InfiniteTalk and LTX LipDub are interesting because they move beyond simple mouth replacement.

InfiniteTalk supports audio-driven video-to-video and image-to-video talking generation. That means it can influence more than lips: head motion, body posture, expression, and longer clip continuity enter the picture. That makes it attractive for generated presenter workflows, but also changes the benchmark. You are no longer asking "did the mouth update cleanly?" You are asking whether a larger video model preserved identity, motion, framing, and speech alignment together.

LTX LipDub is similarly ecosystem-shaped. It is an IC-LoRA trained on top of LTX 2.3 22B for lip dubbing, with Diffusers metadata and a gated community license. That may become useful if the rest of your video stack is already LTX-shaped, but it is not the same evaluation problem as a local `video + audio -> lip-sync` script.

For both models, the next benchmark should separate:

- output quality
- controllability
- hardware cost
- licensing
- repeatability
- whether the model can preserve the original source clip when that is required

The pain-point split matters here. InfiniteTalk can be attractive because it regenerates broader talking-video motion, but users also report slow inference, 720p memory pressure, and degradation outside the face. LTX LipDub is promising for short redubbing, but its early beta model card and launch discussions point to clip-length, language, teeth sharpness, and workflow constraints that make it watchlist material rather than a default.

Older baselines still have a role

VideoReTalking and Wav2Lip should not be the first models I would promote today, but they still matter as baselines.

Wav2Lip is the classic mental model: input video, input audio, synchronized mouth region. It is easy to understand, widely wrapped, and useful for sanity checks. Its weakness is exactly why newer systems exist: visual fidelity, texture consistency, mouth patch artifacts, and aging dependencies.

VideoReTalking is more structured. It decomposes talking-head editing into expression normalization, audio-driven lip sync, and face enhancement. That multi-stage design is useful to study, even if the production tradeoff is heavier preprocessing and older model quality.

Diff2Lip belongs in the research-reference bucket. Diffusion-based lip sync is directionally relevant, but production adoption depends on whether there is a maintained, repeatable repo that can pass the same fixtures.

For older baselines, issue trackers are most useful as failure checklists. Wav2Lip still draws mouth-box, face-detection, checkpoint, audio, and non-commercial-license complaints. VideoReTalking issues lean toward older-stack friction such as CMake, dlib, GPEN, Docker, pixelated eyes, mouth shadows, and sync drift after a few seconds.

Portrait animation is adjacent, not equivalent

Hallo, Hallo2, and EchoMimic are useful when the input is a still portrait or a generated presenter reference. They are not direct replacements for existing-video dubbing.

This distinction matters. If your product needs to dub a founder clip, a sales video, or a real customer testimonial, preserving the source video is a requirement. If your product needs to create a presenter from a headshot, portrait animation models are in scope.

Mixing those categories makes benchmarks look more impressive and less useful.

How I would run the next bakeoff

Use a small fixture set before making any production routing change:

Fixture	Why it matters
Clean front-facing clip	Basic viseme timing and mouth shape.
Off-axis face	Checks whether the model handles real social-video angles.
Fast speech with plosives	Exposes timing drift and mushy mouth shapes.
Accented or non-English speech	Tests whether audio features generalize to the product audience.
Vertical compressed social clip	Catches crop, blur, and square-workflow problems.
Clip with partial occlusion	Tests hands, microphones, subtitles, hair, and product overlays.

For every candidate, capture:

- source video and source audio
- model version and checkpoint
- preprocessing and crop command
- hardware or provider
- runtime
- estimated cost or credits when using an API
- final output

- visible failure modes
- full-frame preservation notes
- teeth and mouth-edge artifact notes
- input validation notes for face angle, cuts, speaker count, and occlusion
- moderation, watermark, and retry behavior when using a hosted provider
- human review notes
- license and model-weight terms

The promotion rule should be strict: a candidate replaces LatentSync 1.6 only if it beats LatentSync on the same fixtures for mouth fidelity, temporal sync, identity preservation, and preprocessing burden without introducing worse production constraints.

Practical recommendations

If you need a production-oriented shortlist today:

1. Start with LatentSync 1.6 as the local baseline.
2. Retest KeySync as the closest direct alternative.
3. Keep MuseTalk for speed-constrained preview or fallback experiments.
4. Treat InfiniteTalk as a heavier talking-video candidate, not a simple lip-sync swap.
5. Track LTX LipDub if your video stack is already moving toward LTX 2.3.
6. Keep Wav2Lip and VideoReTalking as interpretability and sanity baselines.
7. Use Hallo and EchoMimic when the input is a portrait, not when the requirement is to preserve an existing video.

The main lesson is operational: do not pick a lip-sync model from demo clips alone. Pick it from the failure modes your product will actually hit.

Sources

- LatentSync GitHub - accessed May 22, 2026.
- LatentSync issue #67, output blur - accessed May 22, 2026.
- KeySync GitHub - accessed May 22, 2026.
- KeySync issue #35, identity and teeth - accessed May 22, 2026.
- KeySync issue #29, color and FPS or sample-rate assumptions - accessed May 22, 2026.
- MuseTalk GitHub - accessed May 22, 2026.

- MuseTalk issue #140, out-of-sync output - accessed May 22, 2026.
- InfiniteTalk GitHub - accessed May 22, 2026.
- InfiniteTalk issue #156, source-video degradation - accessed May 22, 2026.
- InfiniteTalk issue #197, runtime and GPU pressure - accessed May 22, 2026.
- LTX 2.3 LipDub model card - accessed May 22, 2026.
- LTX LipDub beta discussion on Reddit - accessed May 22, 2026.
- VideoReTalking GitHub - accessed May 22, 2026.
- Wav2Lip GitHub - accessed May 22, 2026.
- Wav2Lip issue #415, mouth box artifact - accessed May 22, 2026.
- Diff2Lip paper - accessed May 22, 2026.
- Hallo GitHub - accessed May 22, 2026.
- EchoMimic GitHub - accessed May 22, 2026.
- TalkVerse paper - accessed May 22, 2026.
- Sync Labs troubleshooting docs - accessed May 22, 2026.
- Sync Labs lip sync quality docs - accessed May 22, 2026.
- Runway Act-Two docs - accessed May 22, 2026.
- Runway lip sync guide and credit example - accessed May 22, 2026.
- HeyGen video translation help center - accessed May 22, 2026.
- HeyGen localization support complaint on Reddit - accessed May 22, 2026.
- HeyGen Avatar III feature-availability complaint on Reddit - accessed May 22, 2026.
- HeyGen Product Hunt reviews - accessed May 22, 2026.
- D-ID FAQ - accessed May 22, 2026.
- Reddit lip sync API cost thread - accessed May 22, 2026.
- Internal Instavar lip sync forum sentiment report, May 22, 2026.