PLAYBOOK 02 OF 3

Hybrid AI UGC Video Creation

Best of both worlds. Run free open-source tools locally for unlimited output, use cloud only for premium features. $45-80/month for unlimited videos after initial GPU investment.

Cost: $45-80/month ongoing (after GPU purchase)

GPU Investment: $250-1,500 one-time

Output: Unlimited videos per month

Built by Shivendra Rawat | Founder, Second Step
getsecondstep.com | growth@shivendrarawat.com

Local vs. Cloud: Decision Matrix

Not everything should run locally. Some tasks are better (and cheaper) in the cloud. Here is how to decide.

TASK	RUN LOCALLY?	RUN CLOUD?	WHY
Voice/TTS	**Yes**	Backup only	Chatterbox TTS is free and matches ElevenLabs quality
Lip-sync	**Yes**	For complex scenes	MuseTalk runs well on RTX 3060+, free
4K Upscaling	**Yes**	No	Real-ESRGAN is free, fast, and excellent quality
Transcription	**Yes**	No	Whisper is free and highly accurate
Script Writing	No	**Yes**	Cloud LLMs (Claude/GPT) are better than local models for ad copy
Premium Avatars	No	**Yes**	HeyGen's avatar quality is hard to replicate locally
Video Assembly	Either	Either	FFmpeg locally or Creatomate cloud -- both work

> **The hybrid advantage:** By running TTS, lip-sync, upscaling, and transcription locally, you eliminate 60-70% of cloud costs. The remaining 30-40% (premium avatars, script writing) is where cloud tools genuinely add value.

GPU Guide

Your GPU determines what you can run and how fast. Here are realistic options:

GPU	VRAM	PRICE (USED)	WHAT IT RUNS	BEST FOR

RTX 3060	12GB	$200-250	Most models comfortably	Best value starter
RTX 3080	10-12GB	$350-450	Faster inference, parallel tasks	Serious production
RTX 4070 Ti	12GB	$500-600	All models, fast	Professional workflow
RTX 4090	24GB	$1,200-1,500	Everything, simultaneously	Agency-scale production

> **Our recommendation:** Start with a used RTX 3060 12GB ($250). It handles Chatterbox TTS, MuseTalk, Real-ESRGAN, and Whisper without issues. Upgrade only when render times become a bottleneck.

Local Tools: Setup Guide

CHATTERBOX TTS (VOICE GENERATION)

What it is: Open-source text-to-speech engine that produces remarkably natural speech. In blind tests, it matches or beats ElevenLabs for most voices.

Cost: Free (open-source)

VRAM Required: 4-6GB

Setup:

1. Install Python 3.10+ and CUDA toolkit
2. Clone the Chatterbox repository from GitHub
3. Install dependencies: `pip install -r requirements.txt`
4. Download the model weights (auto-downloads on first run)
5. Run with your script text as input

Pro tip: Feed it a 30-second voice sample to clone any voice style. The cloning quality is excellent for TTS purposes.

MUSETALK (LIP-SYNC)

What it is: Real-time lip-sync model that generates natural mouth movements from audio input. Works with any face image or video.

Cost: Free (open-source)

VRAM Required: 6-8GB

Setup:

1. Clone the MuseTalk repository
2. Install dependencies (PyTorch, face detection models)
3. Download pretrained weights
4. Input: source face image/video + audio file
5. Output: lip-synced video

Quality note: MuseTalk produces good results for straight-to-camera talking head content. For complex angles or multiple speakers, HeyGen's cloud solution is still superior.

REAL-ESRGAN (4K UPSCALING)

What it is: AI upscaling model that takes 720p or 1080p video and upscales to crisp 4K. Also removes compression artifacts and improves detail.

Cost: Free (open-source)

VRAM Required: 4-6GB

Speed: ~2-5 seconds per frame on RTX 3060

When to use: After assembling your final video. Upscale the complete output rather than individual clips to save time.

WHISPER (TRANSCRIPTION + CAPTIONS)

What it is: OpenAI's speech-to-text model. Generates accurate transcriptions with timestamps -- perfect for auto-generating captions.

Cost: Free (open-source)

VRAM Required: 2-4GB (medium model)

Use case: Transcribe your AI-generated voiceover, then use the timestamps to burn in captions with FFmpeg or your editing tool.

Cost Comparison: Local vs. Hybrid vs. Cloud

	100% CLOUD	HYBRID (RECOMMENDED)	100% LOCAL
Monthly Cost	$89-150/mo	$45-80/mo	$20/mo (LLM only)
Upfront Cost	$0	$250-500 (GPU)	$500-1,500 (GPU)
Videos/Month	15-30 (credit limited)	Unlimited	Unlimited
Avatar Quality	Excellent	Good-Excellent	Good
Setup Time	30 minutes	4-8 hours	1-2 days
Technical Skill	Low	Medium	High
Best For	Getting started fast	Regular production	Technical teams

The Hybrid Workflow

Phase 1: Script (Cloud)

Use Claude or ChatGPT to generate ad scripts. Cloud LLMs are still significantly better than local models for persuasive, direct-response copywriting.

Phase 2: Voice (Local)

Run Chatterbox TTS locally with your target voice profile. Generate the voiceover in seconds. Free, unlimited, no credit limits.

Phase 3: Avatar (Cloud or Local)

For premium quality: Use HeyGen standard avatars (cloud). For unlimited output: Use MuseTalk with a stock presenter image (local). We recommend HeyGen for client-facing content and MuseTalk for A/B test variants.

Phase 4: Assembly (Local)

Use FFmpeg or CapCut to combine avatar video, captions (from Whisper), product shots, and branding. This step is free regardless of approach.

Phase 5: Upscale (Local)

Run Real-ESRGAN to upscale your final output to 4K. This step alone saves ~$5-10 per video compared to cloud upscaling services.

Phase 6: QC and Export

Final quality check and export in platform-specific formats.

> **The payoff:** After the initial GPU investment ($250 for a used RTX 3060), your per-video cost drops to under $2 for most content. The GPU pays for itself within the first month of production.

Getting Started Checklist

HARDWARE

- NVIDIA GPU with 12GB+ VRAM (RTX 3060 minimum)
- 16GB+ system RAM
- 50GB+ free SSD storage for models
- Linux or Windows (WSL2 works well)

SOFTWARE STACK

- Python 3.10+
- CUDA Toolkit (matching your GPU driver)
- Chatterbox TTS (voice)
- MuseTalk (lip-sync)
- Real-ESRGAN (upscaling)
- Whisper (transcription)
- FFmpeg (video processing)

CLOUD ACCOUNTS (HYBRID)

- Claude Pro or ChatGPT Plus ($20/mo) for scripts
- HeyGen Creator ($29/mo) for premium avatars
- Optional: Creatomate ($25/mo) for template-based assembly

Need Help Setting Up Your Hybrid Pipeline?

We help teams implement AI UGC production workflows -- from GPU selection to full pipeline setup. Book a free strategy call to discuss your needs.

BOOK YOUR FREE CALL

Second Step | A performance marketing agency powered by AI

getsecondstep.com | growth@shivendrarawat.com

More playbooks at getsecondstep.com/ugc-video-playbooks