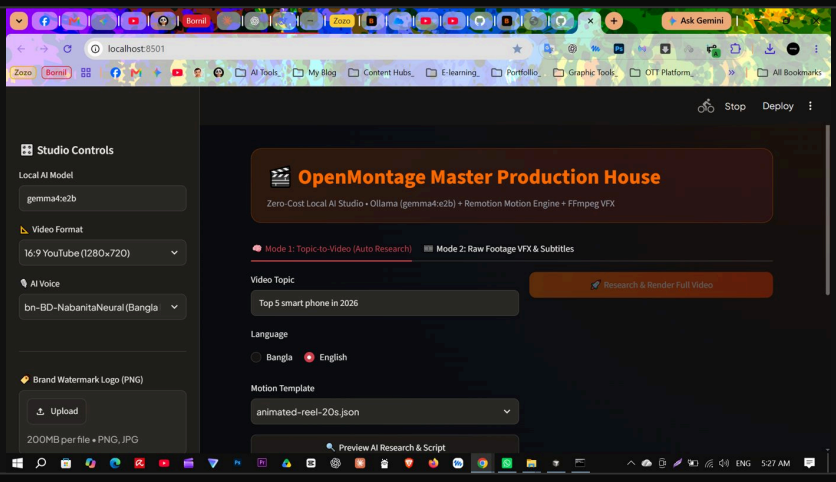
# OpenMontage Details

Open-source, agentic video production system - পুরো বিবরণ, আপনার নিজের লোকাল সেটআপের বিশ্লেষণ, আর visual ও voice lip-sync না মেলার আসল কারণ ও সমাধান।

**12** Pipelines | **100+** Tools | **700+** Skill Files | **60+** Providers | Licence **AGPL-3.0**




আপনার নিজের সেটআপ - OpenMontage Master Production House, localhost:8501। বামে Studio Controls, মাঝে Mode 1: Topic-to-Video। এই ছবিটাই এই গাইডের ৬ নম্বর পৃষ্ঠায় ধাপে ধাপে ব্যাখ্যা করা হয়েছে।

### এটা কী

একটা prompt থেকে শুধু একটা clip নয় - research, script, voice, visual, edit ও render পর্যন্ত পুরো production pipeline।

### আপনার সেটআপ

Ollama (gemma4:e2b) + Remotion + FFmpeg + Microsoft Edge TTS দিয়ে সাজানো zero-cost local studio।

### সবচেয়ে জরুরি অংশ

৭ নম্বর পৃষ্ঠা - কেন YouTube-এর মতো visual আসছে না, আর কেন voice-এর সঙ্গে lip-sync মিলছে না।

01 - পরিচিতি

# OpenMontage একটা clip generator নয়, পুরো production system

OpenMontage নিজেকে বলে বিশ্বের প্রথম open-source, agentic video production system। আপনি plain language-এ বলবেন কী ভিডিও চান - AI agent research, scripting, asset তৈরি, editing আর final composition সামলাবে।

বেশিরভাগ AI video tool একটা prompt থেকে একটা ছোট clip দেয়। OpenMontage তার উল্টো - একটা আসল production team যেভাবে কাজ করে, ঠিক সেই ধাপগুলো agent দিয়ে করায়। তাই এর output শুধু generated footage-এর ওপর নির্ভর করে না; নির্ভর করে কোন pipeline, কোন provider, আর কোন render engine বেছে নেওয়া হয়েছে তার ওপর।

## যে দুই ধরনের ভিডিও এটা বানাতে পারে

**Image-based video**

AI দিয়ে ছবি তৈরি হয়, Piper বা অন্য TTS দিয়ে narration হয়, আর Remotion সেই ছবিগুলোকে animated scene, text card, chart আর word-level caption দিয়ে সাজিয়ে ভিডিও বানায়।

**Real-footage video**

Documentary Montage pipeline free stock ও open archive - Archive.org, NASA, Wikimedia Commons, Pexels, Unsplash - থেকে আসল motion footage খুঁজে, CLIP দিয়ে rank করে, timeline-এ কেটে আসল ভিডিও বানায়। এটা শুধু still ছবি নাড়ানো নয়।

## Reference video থেকে শুরু করা যায়

YouTube video, Short, Reel, TikTok বা নিজের local clip দিলে agent সেটার transcript, pacing, scene, keyframe আর style বিশ্লেষণ করে। তারপর আপনি পান 2-3টি আলাদা concept, কোন tool দিয়ে বানানো হবে তার সৎ ধারণা, শুরুর আগেই cost estimate, আর full production-এর আগে একটা sample।

> **মূল পার্থক্যটা মনে রাখুন:** OpenMontage-এর নিজের কোনো orchestrator code নেই। আপনার AI coding assistant-ই orchestrator। সে YAML pipeline manifest পড়ে, Markdown director skill পড়ে, Python tool ডাকে, নিজে review করে, আর প্রতিটা creative সিদ্ধান্তে আপনার approval নেয়।

## কারা ব্যবহার করতে পারে

Claude Code, Cursor, GitHub Copilot, Windsurf, Codex - যেকোনো AI coding assistant যেটা file পড়তে ও code চালাতে পারে। প্রতিটার জন্য আলাদা instruction file repo-তেই থাকে: `CLAUDE.md`, `CURSOR.md`, `COPILOT.md`, `CODEX.md` - সবগুলোই shared `AGENT_GUIDE.md`-এর দিকে নির্দেশ করে।

02 - কীভাবে কাজ করে

# প্রতিটা ভিডিও একই ৭ ধাপের flow দিয়ে যায়

1 Research > 2 Proposal > 3 Script > 4 Scene Plan > 5 Assets > 6 Edit > 7 Compose

**Research সবার আগে, script তার পরে।** Script-এর একটা শব্দ লেখার আগেই agent live web search করে - YouTube, Reddit, Hacker News, news site ও academic source থেকে data point, audience-এর প্রশ্ন, trending angle আর visual reference জোগাড় করে একটা structured research brief বানায়। তাই ভিডিওর তথ্য guess করা নয়, source-ভিত্তিক।

প্রতিটা ধাপে একটা করে dedicated director skill থাকে - Markdown file, যেখানে agent-কে ঠিক কীভাবে ওই ধাপটা করতে হবে তা শেখানো থাকে। ধাপ শেষে agent নিজে review করে, state checkpoint করে, আর creative সিদ্ধান্তের জায়গায় আপনার approval-এর জন্য থামে।

## Three-Layer Knowledge Architecture

| Layer | কোথায় থাকে | কাজ কী |
|---|---|---|
| **Layer 1** | `tools/` + `pipeline_defs/` | কী কী আছে - executable tool আর pipeline-এর playbook (YAML manifest)। |
| **Layer 2** | `skills/` | কীভাবে ব্যবহার করতে হয় - OpenMontage-এর নিজের convention আর quality bar। |
| **Layer 3** | `.agents/skills/` | ভেতরে কীভাবে কাজ করে - বাইরের technology-র deep knowledge pack। |

## Backlot - Living Storyboard

Chat শুধু বলে agent কী বলেছে। **Backlot** দেখায় production আসলে কী করছে - একটা local board, যেখানে stage জ্বলে ওঠে, script screenplay page হিসেবে আসে, asset তৈরির সময় scene card দেখা যায়, আর প্রতিটা provider decision ও খরচ দেয়ালে থাকে। Asset generation এখানে scene-by-scene contact sheet-এ থামে - take, prompt, per-asset cost, quality score - যাতে render-এর আগেই আপনি visual approve করতে পারেন, পরে নয়।

```
python -m backlot open                # sob project-er library
python -m backlot open <project-id>   # ekta production-er live board
```

> **Render runtime proposal-এর সময়ই lock হয়।** Remotion নাকি HyperFrames - এই সিদ্ধান্ত `render_runtime` হিসেবে lock হয়ে যায়। মাঝপথে চুপচাপ engine বদলে ফেলা OpenMontage-এ governance violation।

03 - PIPELINES

# 12টি pipeline, প্রতিটা আলাদা কাজের জন্য

Pipeline বেছে নেওয়াটাই প্রথম ও সবচেয়ে বড় সিদ্ধান্ত। ভুল pipeline বেছে নিলে বাকি সব ঠিক করলেও output মনমতো হবে না।

| Pipeline | কী তৈরি করে | সেরা ব্যবহার |
|---|---|---|
| **Animated Explainer** | Research, narration, visual ও music সহ AI explainer | Educational content, tutorial, topic breakdown |
| **Animation** | Motion graphics, kinetic typography, animated sequence | Social media, product demo, abstract concept |
| **Avatar Spokesperson** | Avatar দিয়ে presenter video | Corporate comms, training, announcement |
| **Cinematic** | Trailer, teaser, mood-driven edit | Brand film, teaser, promotional content |
| **Clip Factory** | একটা লম্বা source থেকে ranked short clip-এর batch | লম্বা content social media-র জন্য repurpose |
| **Documentary Montage** | Free stock ও open archive থেকে CLIP-indexed আসল footage-এর montage | Video essay, mood piece, paid API ছাড়া real-footage video |
| **Hybrid** | নিজের footage + AI support visual | আছে এমন footage-কে graphics দিয়ে enhance করা |
| **Localization & Dub** | আছে এমন ভিডিওর subtitle, dub ও translation | Multi-language distribution |
| **Podcast Repurpose** | Podcast highlight থেকে video | Podcast marketing, audiogram video |
| **Screen Demo** | Polished software screen recording ও walkthrough | Product demo, tutorial, documentation |
| **Talking Head** | Footage-led speaker video | Presentation, vlog, interview |
| **Character Animation** | SVG rig, pose library ও GSAP timeline দিয়ে cartoon character acting | Cartoon character video, HyperFrames render |

ওপরের তালিকার প্রথম ১১টি README-র pipeline table থেকে নেওয়া; Character Animation README-র free-path ও HyperFrames অংশে আলাদাভাবে বর্ণিত - মোট সংখ্যা অফিসিয়াল বর্ণনায় 12টি pipeline।

04 - TOOLS, PROVIDERS ও FREE PATH

# 100+ tool, 60+ provider - quality আসে provider থেকে

| Category | উল্লেখযোগ্য provider / tool | খেয়াল রাখুন |
|---|---|---|
| **Video Generation** 20+ | Kling, Google Veo 3.1, Runway Gen-4, Seedance, MiniMax, HeyGen, Grok Imagine; local: WAN 2.1/2.2, Hunyuan, CogVideo, LTX | সবচেয়ে ভালো motion quality আসে paid cloud বা local GPU model থেকে। |
| **Image Generation** 15+ | FLUX, Google Imagen 4, GPT Image 2, Recraft, Seedream; local: Stable Diffusion; stock: Pexels, Pixabay, Unsplash | Still ভালো হলেই ভিডিও ভালো হয় না - motion আলাদা ধাপ। |
| **Text-to-Speech** 10+ | ElevenLabs, Google TTS (700+ voice, 50+ language), OpenAI TTS, Azure Speech; local free: Piper | Voice-এর natural timing-ই পরে caption ও lip-sync-এর ভিত্তি। |
| **Music ও Sound** | Suno AI, ElevenLabs Music ও SFX | Music auto-select হয়, তবে style playbook অনুযায়ী। |
| **Lip Sync ও Avatar** | Wav2Lip, SadTalker, MuseTalk, Kling Avatar ও Kling Lip Sync | মুখ নাড়ানো চাইলে এই tool লাগবেই - শুধু TTS দিয়ে lip-sync হয় না। |
| **Analysis** | WhisperX transcription (word-level timestamp), Scene Detect, Frame Sampler, CLIP/BLIP-2 Video Understand | Word-level timestamp-ই সঠিক caption ও sync-এর চাবি। |
| **Post-Production** | FFmpeg, Video Stitch, Video Trimmer, Audio Mixer, Audio Enhance, Color Grade, Subtitle Gen | সবসময় available, সবসময় free। |
| **Composition** | Remotion (React) ও HyperFrames (HTML/CSS/GSAP) | Explainer ও data video-তে Remotion; motion-graphics-heavy কাজে HyperFrames। |

## Zero API Key দিয়েই যা পাওয়া যায়

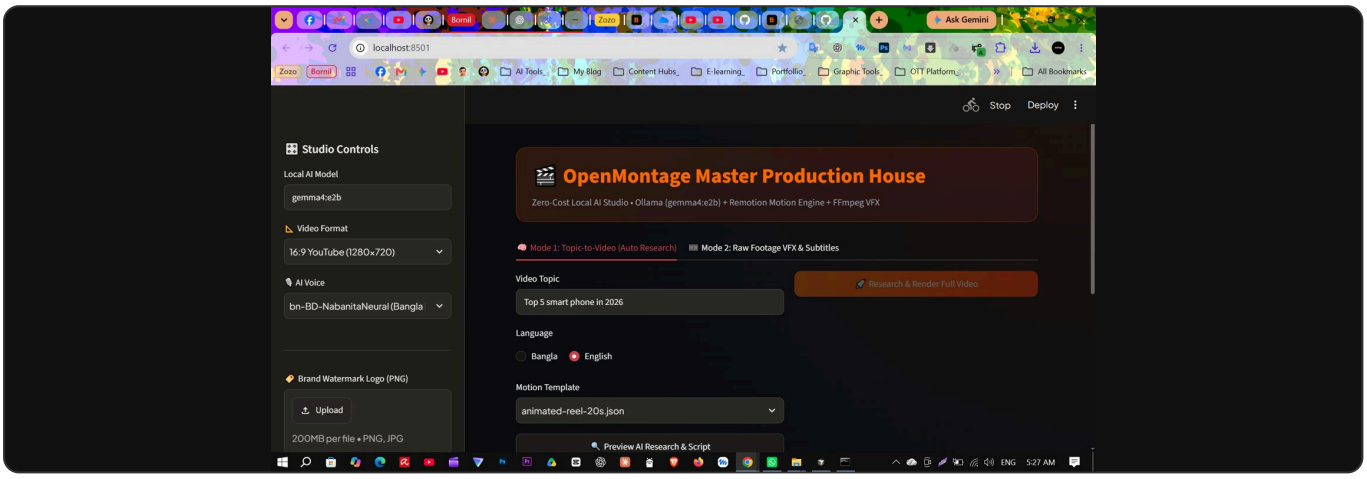
| Capability | Free tool | কী করে |
|---|---|---|
| Narration | Piper TTS | Offline, free text-to-speech narration |
| Open footage | Archive.org + NASA + Wikimedia | Free archival ও documentary footage |
| Extra stock | Pexels + Unsplash + Pixabay | Free stock footage ও image (developer key free) |
| Composition | Remotion / HyperFrames | Animated scene, kinetic typography, word-level caption |
| Post-production | FFmpeg | Encoding, subtitle burn-in, audio mixing, color grading |

খরচের ধারণা (README-র উদাহরণ): image/video provider থাকলে প্রতি short প্রায় $0.15-$1.50; full setup-এ $1-$3। উদাহরণ - Pixar-style short "The Last Banana" $1.33, transformation film প্রায় $4, model showcase প্রায় $5। Default budget cap $10, আর $0.50-এর বেশি একক খরচে approval লাগে।

05 - আপনার নিজের সেটআপ

# অফিসিয়াল OpenMontage বনাম আপনার localhost panel

**সবচেয়ে জরুরি সত্য:** অফিসিয়াল OpenMontage-এ কোনো Streamlit dashboard নেই। ওটা agent-first - Claude Code / Cursor-এর মতো assistant-কে prompt দিলেই pipeline চলে। আপনার `localhost:8501` -এর "OpenMontage Master Production House" panel-টা একটা custom local wrapper - GitHub থেকে প্রয়োজনীয় software install করে আপনার নিজের কাজের জন্য সাজানো। তাই YouTube-এ দেখা demo-র সঙ্গে আপনার panel-এর behaviour হুবহু মিলবে না, কারণ ভেতরের tool-গুলো আলাদা।



আপনার panel: বামে Studio Controls, মাঝে OpenMontage Master Production House - Zero-Cost Local AI Studio।

| আপনার panel-এর option | এখন যা set আছে | মানে কী |
|---|---|---|
| Local AI Model | `gemma4:e2b` (Ollama) | Script ও research লেখার local model। e2b মানে ছোট model - দ্রুত, কিন্তু বড় cloud model-এর মতো গভীর script ও scene timing আশা করা যায় না। |
| Video Format | 16:9 YouTube (1280x720) | Landscape output, 720p। YouTube-এর অফিসিয়াল profile হলো 1920x1080 - আপনারটা তার অর্ধেক resolution। |
| AI Voice | `bn-BD-NabanitaNeural` (Bangla, Edge TTS) | Microsoft Edge TTS-এর বাংলা voice। এটি OpenMontage-এর অফিসিয়াল TTS তালিকার (Piper / ElevenLabs / Google / OpenAI) বাইরের, আপনার wrapper-এর নিজের সংযোজন। |
| Language | English selected | **এখানেই mismatch:** voice Bangla, কিন্তু language English - ৭ নম্বর পৃষ্ঠায় বিস্তারিত। |
| Video Topic | Top 5 smart phone in 2026 | Mode 1-এ topic দিলেই research ও script auto তৈরি হয়। |
| Motion Template | `animated-reel-20s.json` | Fixed 20-second template। Narration 20 second-এর চেয়ে লম্বা বা ছোট হলেই sync ভাঙবে। |
| Brand Watermark | PNG upload (200MB পর্যন্ত) | আপনার logo watermark হিসেবে ভিডিওতে বসে। |

**Mode 1 - Topic-to-Video (Auto Research):** topic দিলে script, voice, visual ও render একসঙ্গে হয়। **Mode 2 - Raw Footage VFX & Subtitles:** নিজের তোলা footage-এ VFX ও subtitle যোগ করার mode। দুটো mode-এর pipeline আলাদা - Mode 2-তে lip-sync আশা করা, আর Mode 1-এ মুখের ভিডিও আশা করা, দুটোই ভুল প্রত্যাশা।

06 - সমস্যার আসল কারণ

# Visual মনমতো হচ্ছে না, voice lip-sync মিলছে না - ৫টি কারণ

এটা একটা bug নয়। পাঁচটা আলাদা কারণ একসঙ্গে কাজ করছে - কোনোটা setting-এর, কোনোটা tool বাছাইয়ের, কোনোটা প্রত্যাশার।

1 **Language আর Voice-এর mismatch - সবচেয়ে বড় কারণ।** আপনার screenshot-এ Language = English, কিন্তু AI Voice = `bn-BD-NabanitaNeural` (Bangla)। Bangla voice দিয়ে English script পড়ালে উচ্চারণ, গতি আর word timing - সব এলোমেলো হয়। Caption ও scene timing সেই ভুল timing-এর ওপর তৈরি হয়, তাই চোখে sync ভাঙা মনে হয়। নিয়ম একটাই: Bangla voice হলে Language-ও Bangla, English voice হলে Language-ও English।

2 **আপনার pipeline-এ আসল lip-sync tool-ই নেই।** Lip-sync মানে audio শুনে মুখের নড়াচড়া তৈরি করা। OpenMontage-এ এর জন্য আলাদা tool আছে - Wav2Lip, SadTalker, MuseTalk, Kling Lip Sync - আর সেগুলো Talking Head / Avatar pipeline-এ চলে। Topic-to-Video-তে শুধু TTS voice-এর ওপর ছবি বা template বসালে সেটা narration + visual, lip-sync নয়। মুখই যদি generated character না হয়, মেলার মতো lip থাকেই না।

3 **Fixed template বনাম narration-এর দৈর্ঘ্য।** `animated-reel-20s.json` template 20 second ধরে বানানো। আপনার narration যদি 26 second হয়, শেষ 6 second visual আগেই শেষ; 14 second হলে শেষে ফাঁকা বা টানা অংশ থাকে। সঠিক পদ্ধতি হলো WhisperX দিয়ে word-level timestamp নিয়ে প্রতিটা scene-এর duration narration অনুযায়ী সাজানো - Remotion-এর word-level caption ঠিক এই কাজটাই করে।

4 **YouTube demo আর আপনার setup-এর provider এক নয়।** README-র showcase ভিডিওগুলো Veo, Kling v3, Google Chirp3-HD-এর মতো paid provider দিয়ে বানানো, খরচ $1.33 থেকে প্রায় $5 পর্যন্ত। আপনার zero-cost local path - ছোট local model, Edge TTS, still image ও Remotion template - দিয়ে সেই motion quality আসবে না। এটা setup-এর দোষ নয়, tier-এর পার্থক্য।

5 **ছোট local model-এর script ও timing।** `gemma4:e2b` দ্রুত local model, কিন্তু scene ভাগ করা, প্রতিটা scene-এ কত second narration যাবে, hook কোথায় - এসব judgment-এ বড় model-এর চেয়ে দুর্বল। Script-এর timing দুর্বল হলে edit-ও দুর্বল হয়, যত ভালো renderer-ই থাকুক।

> **সারকথা:** sync-এর সমস্যা setting ঠিক করলেই অনেকটা মিটবে (কারণ 1 ও 3)। Visual-এর সমস্যা মিটবে provider বদলালে (কারণ 4) - হয় free real-footage path, নয় GPU/paid video model। দুটো সমস্যার সমাধান এক জায়গায় নেই।

07 - সমাধান ও সঠিক ব্যবহার

# ঠিক করার checklist

## Sync ঠিক করতে - আগে এগুলো

- **Language = Voice-এর ভাষা করুন।** Bangla ভিডিওতে: Language Bangla + `bn-BD-NabanitaNeural`। English ভিডিওতে: Language English + একটা English Edge voice। দুটো কখনো মেশাবেন না।
- **আগে শুধু voice render করে শুনুন।** Voice ঠিকঠাক, স্বাভাবিক গতিতে পড়ছে কিনা নিশ্চিত হয়ে তারপর visual render করুন।
- **Template duration narration অনুযায়ী বদলান।** Fixed 20s template-এ লম্বা script দেবেন না - script ছোট করুন, নয়তো template-এর duration বাড়ান।
- **মুখের ভিডিও চাইলে pipeline বদলান।** Talking Head / Avatar pipeline + Wav2Lip বা MuseTalk ছাড়া lip-sync হবে না - Topic-to-Video দিয়ে চেষ্টা করে সময় নষ্ট করবেন না।
- **Caption word-level timing থেকে নিন।** WhisperX / Remotion-এর word-level caption ব্যবহার করুন।

## Visual ভালো করতে - free-first পথ

- **Free পথে সেরা visual:** Documentary Montage pipeline - "use real footage only" লিখে prompt দিন। Archive.org, NASA, Wikimedia, Pexels থেকে আসল motion footage এলে ভিডিও আর slideshow মনে হবে না।
- **Resolution বাড়ান:** 1280x720-এর বদলে YouTube profile 1920x1080 ব্যবহার করুন।
- **Render-এর আগে approve করুন:** Backlot-এর contact sheet-এ scene দেখে তবেই full render - খারাপ visual-এ পুরো render নষ্ট হবে না।
- **Paid/GPU তখনই:** free-first নিয়ম অনুযায়ী paid provider তখনই ভাবুন, যখন free পথে নির্দিষ্ট কাজ একেবারেই হচ্ছে না (প্রতি short প্রায় $0.15-$3)।

## Platform output, style ও governance

| Profile | Resolution | Profile | Resolution |
|---|---|---|---|
| YouTube Landscape | 1920x1080 (16:9) | YouTube Shorts / Reels / TikTok | 1080x1920 (9:16) |
| YouTube 4K | 3840x2160 (16:9) | Instagram Feed | 1080x1080 (1:1) |
| LinkedIn | 1920x1080 (16:9) | Cinematic | 2560x1080 (21:9) |

Style playbook: Clean Professional (corporate / educational), Flat Motion Graphics (social / TikTok), Minimalist Diagram (technical)। Quality gate: proposal, script, scene plan, asset ও publish - প্রতিটা ধাপে human approval লাগে। Render-এর পরেও ffprobe, frame, audio ও subtitle check হয়। Provider বাছাই হয় 7 dimension-এর score দিয়ে: task fit 30%, quality 20%, control 15%, reliability 15%, cost 10%, latency 5%, continuity 5%।

## Quick start (অফিসিয়াল)

```
Prerequisite: Python 3.10+  |  FFmpeg  |  Node.js 18+  |  AI coding assistant
git clone https://github.com/calesthio/OpenMontage.git ; cd OpenMontage ; make setup
```

Licence: AGPL-3.0 - free-তে ব্যবহার করা যায়, তবে modified version network service হিসেবে চালালে তার source প্রকাশ করতে হয়; paid product-এর ভেতরে ব্যবহারের আগে শর্তটা মাথায় রাখুন। সূত্র: OpenMontage GitHub README (calesthio/OpenMontage), 16 November 2026। আপনার panel-এর তথ্য আপনার নিজের screenshot থেকে নেওয়া।