THE SOLO ENTREPRENEUR

FREE · LOCAL · STEP BY STEP

The AI Influencer Playbook

How we built a realistic AI creator that scripts, speaks and edits her own reels on one Mac, with free, open models and no per-video fees.



The full system, with the real prompts, settings, mistakes and fixes
From 29 finished reels made in November 2026

Cover: our AI host, generated on a Mac. Not a real person.

INSIDE THIS PLAYBOOK

What you'll build, and how

By the end you'll have a repeatable system that turns a script into a finished, captioned, vertical reel, with a consistent AI host who looks and sounds the same in every video.

0 **Our story & the numbers**
What we built and what it took

✦ **The ground rules**
Consent, disclosure, honesty

1 **Set up your studio**
Hardware, tools, models, install

2 **Design the face**
One face, many outfits, real places

3 **Give her a voice**
Cloning with consent, pacing

4 **Write reels that work**
Formats, hooks, on-camera time

5 **Make her talk**
Audio-to-video talking shots

6 **Pick the best take**
Mouth strips and rejection rules

7 **Edit like a pro**
Graphics, real sources, screen steps

8 **Mix, check, ship**
Loudness, captions, QA

9 **Run it every week**
Batch schedule and timings

✚ **Fix-it guide & checklists**
Every problem we hit, solved

HOW TO READ THIS

Each chapter gives the **why** in a sentence or two, then the exact **steps**, then the **mistakes** we made so you can skip them. Commands are shown for a Mac with Apple silicon. Copy them as they are, then change the names and prompts to fit your own creator.

CHAPTER 0 · OUR STORY

Eight days, one Mac, twenty-nine reels

We wanted an educational creator who could post AI-tool and startup reels for founders every day, without a camera, a studio or a monthly avatar subscription. So we built one, entirely on a Mac mini, with open models.

29	*81*	*8*	*₹0*
finished reels	versions rendered	days, start to finish	software cost per reel

How it actually went

DAY	WHAT HAPPENED
Day 1	First local video tests. Learned that short shots (3–4 s) hold a face, and long ones drift.
Day 2	Rejected our first host (she looked too old). Designed a new face, cloned a friend's voice with her permission, made the first edited reels.
Day 3	Switched to the newer LTX-2.5 video model after a side-by-side test. Made six reels overnight.
Day 3, morning	Verdict: "the lip sync sucks." Found the cause (her starting photos had open grins) and remade all six with calm, closed-lip stills.
Day 4	She "looked too perfect to be human." Tested real places and real light: cafés, couches, balconies.
Day 5	Built the first full tutorial reel with real screen recordings. Slowed the voice down and added proper pauses.
Days 6–8	Tried cloning a real face too, then produced a full series of tutorial and news reels.

THE ONE BIG LESSON

Every quality jump came from **fixing an input** (the starting photo, the voice pacing, the prompt), not from more computing power. Most of this book is about getting those inputs right.

THE SYSTEM ON ONE PAGE

From idea to reel in eight stages

Every reel goes through the same pipeline. Stages 1–3 you do once; stages 4–8 you repeat for every video.

1 ONCE

Studio

Install the free video, image and voice models on your Mac.

2 ONCE

Face

Design one face, then make outfits and places from it.

3 ONCE

Voice

Clone a voice you have consent to use.

4

Script

Hook, story or tutorial, lesson, comment CTA. 30–55 s.

5

Talking shots

2–3 short lines on camera, animated from her photo and the voice.

6

Pick takes

Two takes per line; keep the one with the clearest mouth.

7

Edit

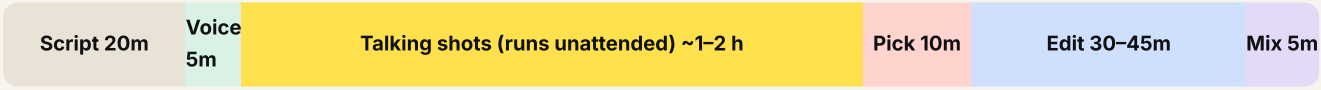
Graphics, real sources, screen steps, captions, cuts.

8

Mix & check

Music, effects, loudness, transcript and frame check.

Where the time goes (one 35-second reel)



Your hands-on time: about 75–90 minutes

Machine time: 1–2 hours, ideal overnight

WHAT'S FREE

Every model and tool in this book is free and open, and runs offline once downloaded. Your costs are the Mac, electricity and your time.

WHAT IT WON'T DO

It isn't instant. A talking shot takes 7–40 minutes to render. Plan reels a day ahead and let the Mac work overnight.

BEFORE YOU BUILD

The ground rules

An AI creator is powerful because people trust faces. Earn that trust honestly or the whole channel is at risk. These rules aren't optional.

1 · SAY SHE'S AI

Use each platform's AI label on every post, and say it plainly in the bio. Never present her as a real person.

2 · CONSENT FOR EVERY VOICE

Only clone a voice whose owner has given you clear, written permission. Never clone a celebrity, a client or a stranger.

3 · INVENT THE FACE

Generate an original face. Never use or "improve" a real person's photo without their consent.

4 · CLEARLY ADULT, NEVER SEXUALISED

Keep the character clearly adult and the styling tasteful, even when you go for fashionable looks.

5 · NO FAKE EXPERIENCE

She can explain and report. She can't say "I tried this and made ₹1 lakh." No invented testimonials.

6 · SHOW YOUR SOURCES

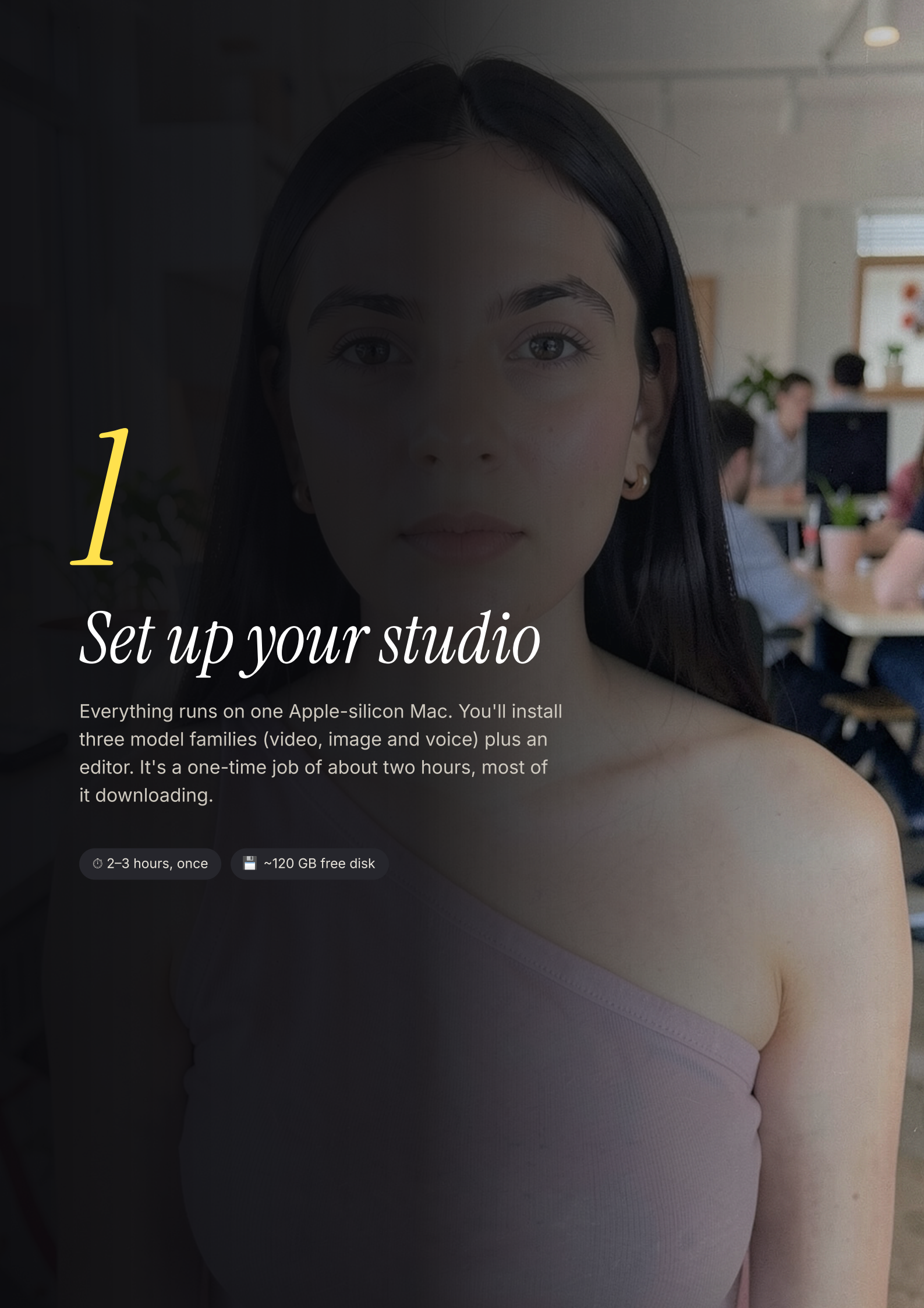
Every number on screen comes from a real page, shown with its name and date. If you can't source it, cut it.

7 · TEACH SAFE TUTORIALS

For outreach tutorials, teach drafts and review before sending, an unsubscribe line, and local anti-spam law.

8 · LABEL RECREATIONS

If a screen is a recreation, not a real recording, tag it on screen ("demo recreation, sample data").



1

Set up your studio

Everything runs on one Apple-silicon Mac. You'll install three model families (video, image and voice) plus an editor. It's a one-time job of about two hours, most of it downloading.

⏱ 2–3 hours, once 💾 ~120 GB free disk

What you need

Hardware

PART	MINIMUM	WHAT WE USED
Mac	Apple silicon (M-series), 32 GB memory	Mac mini, 32 GB
Disk	120 GB free (models are big)	Video model alone is 42 GB
Power	Plugged in, sleep disabled during renders	Renders ran 2–4× slower when the Mac throttled

The free tool stack

JOB	TOOL	MODEL	SIZE
Talking video	`ltx-2-mlx` 0.18	LTX-2.5, 8-bit, distilled	42 GB
Photos of her	`mflux` 0.24	FLUX.2 Klein 4B, 8-bit	~8 GB
Voice clone	`mlx-audio` 0.6.4	VibeVoice 1.5B	~6 GB
Transcripts	`mlx-audio`	Parakeet TDT 0.6B v3	~2 GB
Editing	Remotion (React)	None. Your edit is code	small
Screen recording	Playwright + Chromium	None	small
Audio/video glue	`ffmpeg`	None	small
Take checks	MediaPipe face landmarks	Face Landmarker	small

WHY THESE MODELS

We tested LTX-2.3 against LTX-2.5 on the same line and the same photo, and picked 2.5 by eye. FLUX.2 Klein 4B makes a photo in about a minute and, crucially, can **edit** one photo while keeping the face. That's how one face becomes a whole wardrobe.

MEMORY TIP

Never run the photo model and the video model at the same time. On 32 GB they fight over the GPU and one of them times out.

Install, step by step

1. **Developer tools, Homebrew and helpers.** Open Terminal and run:

```
xcode-select --install
/bin/bash -c "$(curl -fsSL
https://raw.githubusercontent.com/Homebrew/install/HEAD/install.sh)"
brew install uv ffmpeg node
```

2. **The video model code (LTX-2 for Apple silicon).**

```
mkdir -p ~/ai-influencer && cd ~/ai-influencer
git clone https://github.com/dgrauet/ltx-2-mlx.git
cd ltx-2-mlx && uv sync --all-extras && cd ..
```

3. **The video weights.** On huggingface.co, open `dgrauet/ltx-2.5-mlx-q8` and accept its terms with your own account. Then sign in on your Mac and download. You only need the **distilled** files (about 42 GB). Skip the large "dev" transformer and LoRA files.

```
uv tool install "huggingface_hub[cli]"
hf auth login            # paste your own token; never share it
hf download dgrauet/ltx-2.5-mlx-q8 --local-dir models/ltx-2.5-q8
```

4. **The photo and voice tools.**

```
uv tool install mflux==0.24.0
uv tool install --prerelease=allow mlx-audio==0.6.4
```

5. **The editor and recorder.**

```
npx create-video@latest edit      # Remotion project for your reels
uv venv rec && rec/bin/pip install playwright mediapipe && rec/bin/playwright install
chromium
```

6. **Test it.** Make one 3-second clip. If it finishes in about 7–10 minutes, you're ready.

```
ltx-2-mlx/.venv/bin/ltx-2-mlx generate -p "A woman smiling in a cafe" --distilled \
  -f 97 --frame-rate 30 -H 1920 -W 1088 --low-ram --model models/ltx-2.5-q8 -o test.mp4
```

KEEP ONE FOLDER PER REEL

We kept everything for a reel together: `voice/`, `shots/`, `stills/`, `rem/` (the edit) and `out/`. When something breaks, you can see exactly which input made it.



2

Design the face

One great base photo is the foundation of everything. Every outfit, every location and every talking shot is **edited from that one image**, which is how she stays the same person in every reel.

⏱ 1 evening 🖼 ~1 min per photo

Generate options, pick one, lock it

Write a description of who she is (age, background, styling, framing and light) and generate eight options across different seeds. Pick the one that feels most like a real person you'd trust. Then **never generate her from text again**: everything after this is an edit of that photo.



Our real option sheet: 8 seeds, two wording variants. We chose option 6 (seed 106).

```
mflux-generate-flux2 --model flux2-klein-4b -q 8 --steps 4 --seed 106 \
  --width 1088 --height 1920 --output face_s106.png --prompt "Candid photo of a beautiful
  23-year-old adult woman, a recent college graduate, with a soft oval face, gentle almond
  eyes and long straight black hair tucked behind one ear. She wears very light natural
  makeup and tiny gold hoop earrings. Tight framing: head, shoulders and upper chest, facing
  the camera at eye level. Bright airy creator studio in soft natural daylight. Candid
  smartphone photo, natural skin texture, sharp focus."
```

MISTAKE WE MADE

Our first host looked older than intended, and viewers notice age instantly. State the age as “23-year-old adult” in the prompt, and look at the photo at phone size before you commit.

One face, a whole wardrobe

To change her outfit, **edit** the base photo and describe only what changes. Start the prompt with a "keep" sentence that pins the face, hair, earrings and framing. We gave every reel its own outfit so the feed doesn't look like one video on repeat.



Six outfits, one face. All edited from the same base photo.

```
mflux-generate-flux2-edit --model flux2-klein-4b -q 8 --steps 4 --seed 111 \
  --image-paths face_s106.png --width 1088 --height 1920 --output sage.png --prompt \
  "Keep this exact same 23-year-old woman: identical face, identical features, same long
  straight black hair, same tiny gold hoop earrings, same fresh look with light natural
  makeup. Tight framing: head, shoulders and upper chest, facing the camera at eye level.
  She now wears a fitted sage green square-neck top. Natural skin texture, candid photo."
```

The calm-still rule (this fixed our lip sync)

The video model pins the **first frame** of every talking shot to your photo. If that photo has an open-mouthed grin, her lips barely move for the first few words, so the start of every shot looks dubbed.

Fix: make a second edit of each outfit with only the expression changed: *"her lips are gently closed and relaxed, resting together with no teeth showing, a calm, neutral, attentive face... as if about to speak."* Use that calm still for every talking shot.

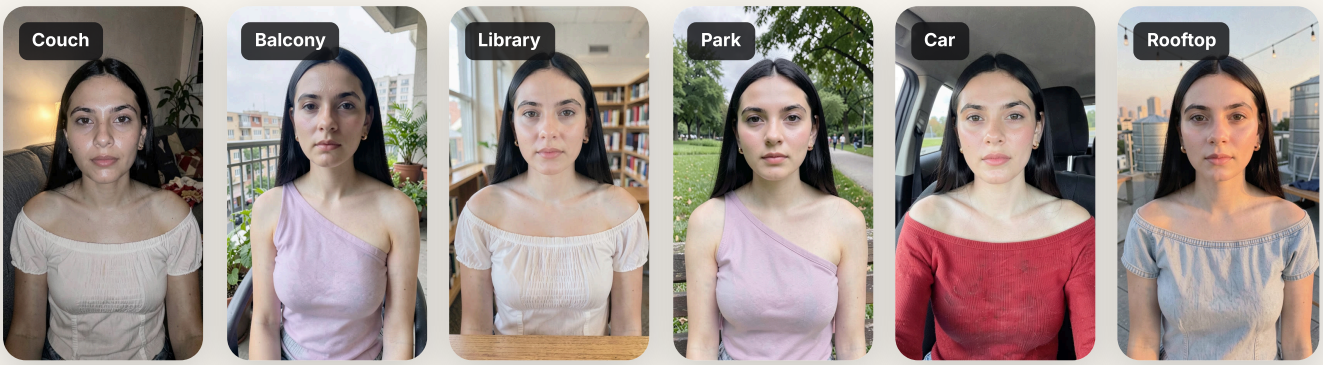


✕ Open grin start

✓ Calm, closed lips

Make her feel real

Studio-perfect light makes an AI face look like an AI face. Putting her in **real, lived-in places** (a café, a couch, a balcony, a car) with natural light is the single biggest realism upgrade. Wide shots are fine, as long as they look real.



All edited from her outfit photos. The café look on the cover became our tutorial host.

Light that works

- Soft window daylight from one side
- Overcast outdoor light
- “The same youthful skin as in this photo”
- “Filmed on a phone propped up in front of her, slight camera grain”

Light that backfired

- Hard side light: she aged 10 years
- Desk lamp + laptop glow: harsh and older
- Asking for “shine” on skin: looked oily
- Hands tucked between legs: melted in video

HANDS RULE

Hands are the first thing to break in video. Either keep them out of frame (waist-up framing) or give them something simple and still to hold, like a coffee mug in both hands. That fixed our worst take.

A voice that's always the same

If each clip invents its own voice, viewers hear a different person every few seconds. So the voice is cloned once, and every line comes from that one clone.

1. **Get consent and a clean sample.** 20–40 seconds of the person speaking naturally, with no music, recorded with written permission. We used a friend's voice with her permission.
2. **Generate the whole script as one take.** Joining separate lines sounds choppy. One continuous take flows like real speech. Make **three takes** with different seeds.
3. **Pick the take by transcript.** Transcribe each take and keep the one whose words match the script best. It catches dropped or swapped words (“send” instead of “save”).
4. **Pace it.** Put a 0.3–0.5 s pause after every sentence and a short breath at commas. Slow a rushed take by 5% with pitch kept (`atempo=0.95`).

```
# voice_takes.py — run with mlx-audio's Python
from mlx_audio.tts.utils import load
import mlx.core as mx
model = load("microsoft/VibeVoice-1.5B", strict=False)
model.lm_head.weight = model.language_model.embed_tokens.weight   # needed for this model
for seed in (6001, 6002, 6003):
    mx.random.seed(seed)
    parts = [r.audio for r in model.generate(SCRIPT, ref_audio="voice_sample.wav", cfg_scale=1.3)]
    # save each take, then transcribe with parakeet-tdt-0.6b-v3 and compare to SCRIPT
```

MISTAKE WE MADE

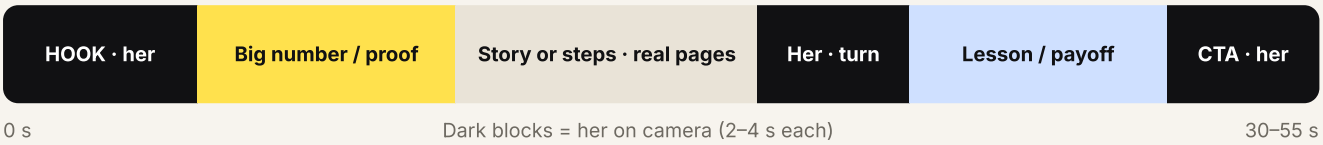
Our cleanup step deleted every pause over 0.09 s. The reel sounded rushed and sentences ran together. Shorten long gaps, but always **keep a real pause** between sentences.

NAMES & BRANDS

Voices stumble on brand names. Respell them phonetically in the script (“Clawd” for Claude) and regenerate just that sentence. Avoid tricky verb forms: “drop the word LEADS in the comments” instead of “comment LEADS”.

She's on camera a quarter of the time

The reels that perform in this niche aren't a talking head for 40 seconds. When we studied top creators frame by frame, they were on camera only **25–33%** of the time. The rest is huge numbers, logos, real pages with highlights and product footage, with a sound on nearly every cut.



Format A · The story reel

A real startup or tool, a surprising number, how it happened, the lesson, a CTA. Example: “One developer built a chatbot tool in two weeks. Today it makes over ten million dollars a year...”

Format B · The tutorial reel

A problem, then numbered steps with real screen recordings, then the payoff. Example: “If you can't find customers for your business, do this. It's completely free.”

Script rules we follow

- Hook in the first sentence, under 3 s
- Every claim has a real source you can show
- She explains, never claims experience
- Short sentences; she pauses after each
- Fast or “smiling” lines go on a card, not her face
- End with one word to comment, and what they get

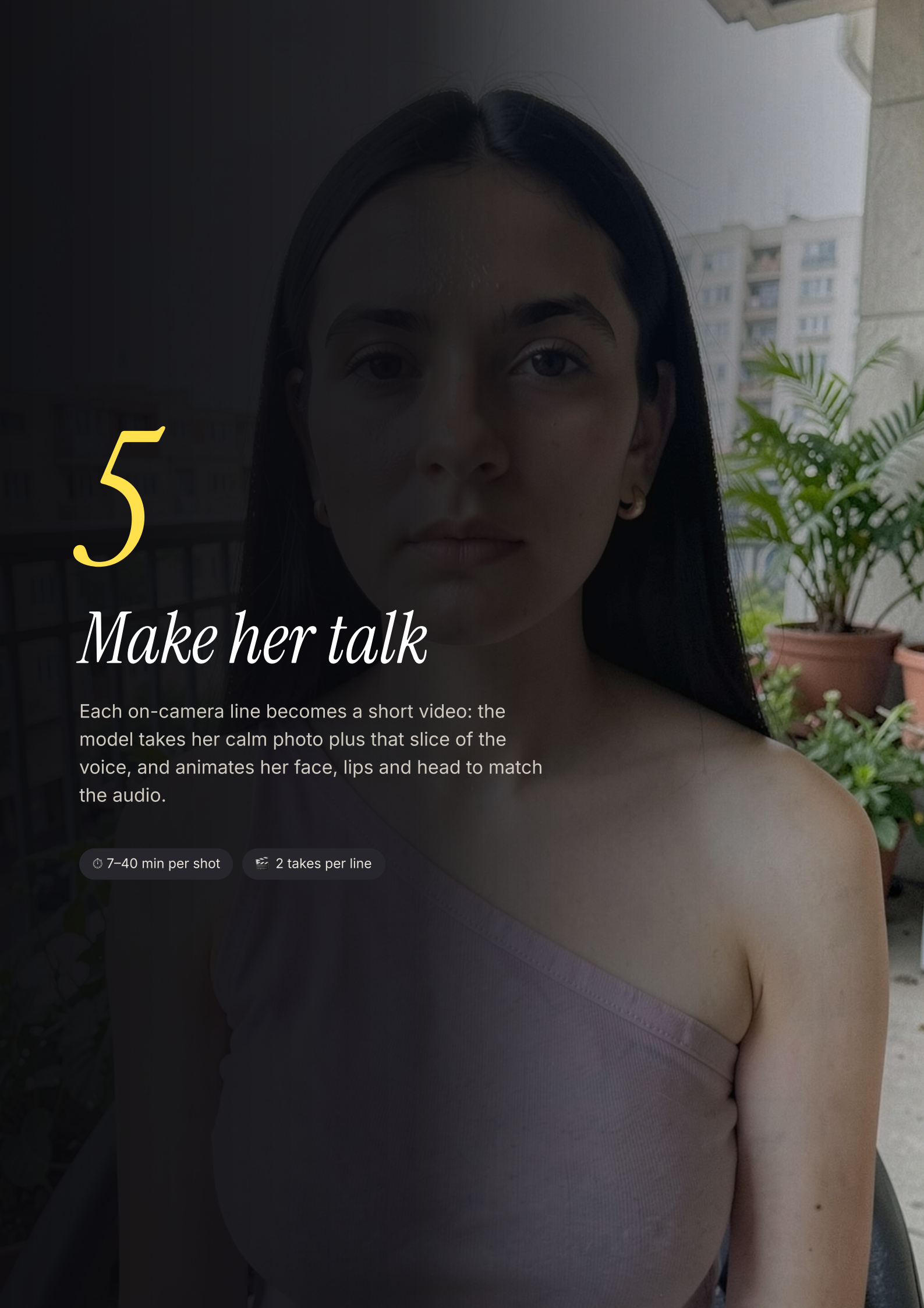
Hooks that stop the scroll

“If you can't find customers for your business, do this. It's free.”

“This automation tool just doubled its value, in under a year.”

“Your next hundred customers are sitting on Google Maps.”

“Why your next phone will cost more, and it's not Apple's fault.”



5

Make her talk

Each on-camera line becomes a short video: the model takes her calm photo plus that slice of the voice, and animates her face, lips and head to match the audio.

⏱ 7–40 min per shot

🎬 2 takes per line

The talking-shot recipe

1. **Cut the line out of the final voice.** Start 0.06 s before her first word and end right on the last word. No trailing silence, or she freezes with her mouth open.
2. **Count frames.** At 30 fps the model needs a frame count of **8n + 1** (65, 73, 97, 113, 121, 145). Round the line's length up to the next one.
3. **Render two takes** with two seeds. About one take in four fails (mouth stuck open, forehead lines, a squint), so you need the spare.

```
ltx-2-mlx a2v --model models/ltx-2.5-q8 --distilled --low-ram \
  -H 1920 -W 1088 -f 97 --frame-rate 30 --seed 971 \
  --audio line_a.wav --image calm.png 0 1.0 -o shots/a_971.mp4 \
  -p "$MOOD $BASE"
```

The prompt that finally worked (from our approved reel)

BASE

```
She is speaking the whole time, naturally, the way a person talks in a relaxed conversation. Her lips and jaw follow every word exactly, opening on the vowels and closing on m, b and p, and she only stops when the sound stops. Her eyebrows and forehead stay still and relaxed, her eyes stay wide enough open and look at the camera lens, blinking only rarely, and her head stays level and almost still. No exaggerated expressions, no wide eyes, no raised eyebrows, no grin, no clenched teeth. Same young woman with the same face, the same outfit, the same place and the same soft light as the first frame, filmed on a phone propped up in front of her.
```

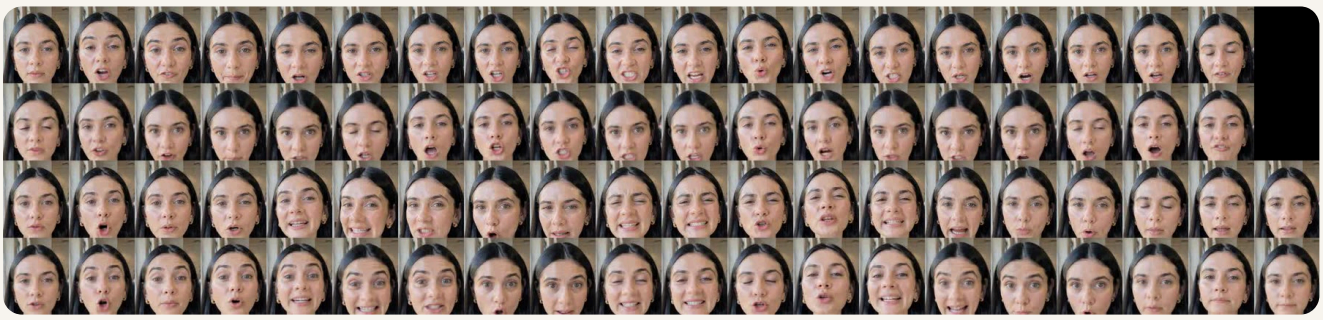
MOOD is one short sentence per line, for example *“She is warm and friendly, inviting the viewer, facing the camera, her head still.”*

LINE LENGTH	FRAMES	RENDER TIME (NORMAL)	ON A SLOW DAY
~1.4 s	41	~7 min	~15 min
~3.2 s	97	~15–18 min	~35 min
~4.8 s	145	~30–40 min	~75 min

Prompt words that break things

Video models take every word literally. These are words we removed after they caused real failures in our takes.

DON'T WRITE	WHAT HAPPENED	WRITE INSTEAD
“selfie”, “phone”	A second phone appeared in her hand in 3 of 4 clips	“filmed on a phone propped up in front of her”, and say where her hands are
“no text, no logos”	Invented fake captions and logos	Leave it out entirely
“jaw drops”, “eyebrows high”	Grimaces, clenched teeth, forehead lines	“no exaggerated expressions, no raised eyebrows”
“nodding”, “leaning in”	Her head dropped out of frame	“her head stays level and almost still”
“smile between phrases”	Grinned through words, lips stopped moving	Calm mood; let the voice carry the warmth
Long shots (6 s +)	Face and earrings drifted, memory ran out	Keep on-camera lines 2–5 s and cut between them



Two takes of the same two lines, frame by frame. Rows 1 and 3 squint and clench their teeth; rows 2 and 4 stay calm. Same photo, same audio, different seed. This is why we always render two.

MISTAKE WE MADE

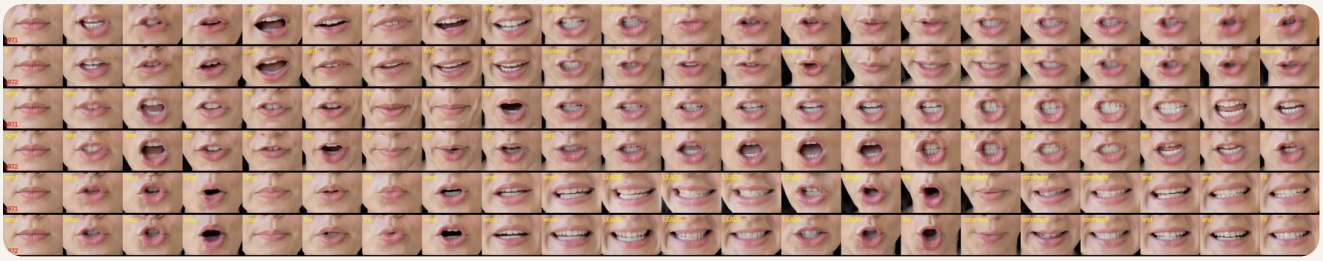
A very fast line said with a smile (“So what does that tell you?”) failed in all ten takes we tried, and she grinned through it every time. Some lines simply belong on a bold title card, with her voice over it.

Judge mouths, not vibes

Watching takes at full speed, everything looks “fine”. So we made two quick checks that turn a take into something you can judge in ten seconds.

Check 1 · The mouth strip

Track the face, crop the mouth every few frames, and label each crop with the word being spoken at that moment. Lips should **close on m, b and p** and **open on vowels**. One take per row, side by side.



Our real mouth strips: three lines, two takes each (top: seed 971, bottom: 972). Look for “O” held across several words, or a frozen grin.

Check 2 · Head jumps & leaning

Track the face position frame to frame and flag sudden jumps. Also track face height: if it grows a lot, she's leaning into the lens. A jump inside the part you use means a cut point or a rejected take.

Reject a take if...

- [ ] her mouth holds one open “O” for several words
- [ ] her lips barely move
- [ ] she grins or clenches her teeth
- [ ] she leans in toward the lens
- [ ] her head jumps inside the used part
- [ ] her eyes close or squint on the first frame

CUT TIMING

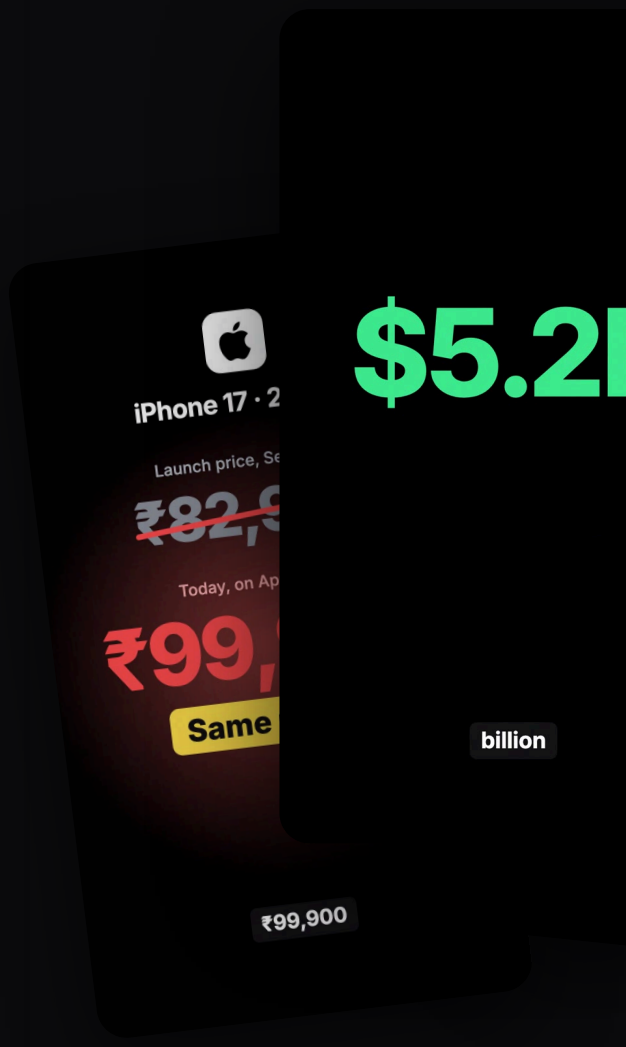
Start each on-camera scene exactly where her shot starts in the voice (scene start minus shot start). Even 0.1–0.3 s early, and the first words look dubbed.

7

Edit like a pro

The edit is where a reel earns attention. We built ours in code with Remotion, so every reel shares the same polished building blocks, and every cut lands exactly on a word.

⏱ 30–45 min per reel 🧩 reusable components



The building blocks

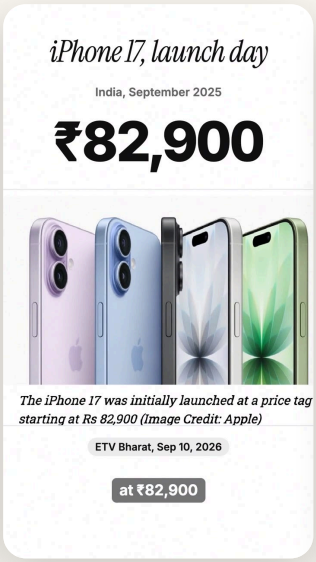
Six scene types cover almost every reel. Each one appears for 1.5–3 s and enters with a quick blur-and-settle, with a sound on the cut.



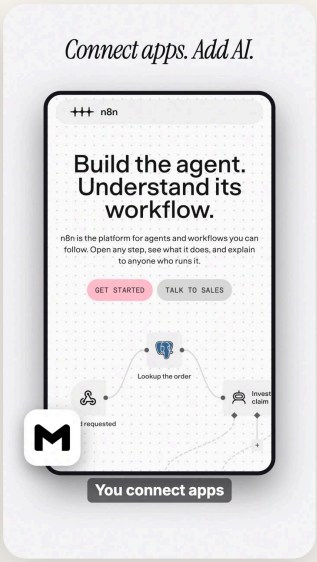
Her, with a headline. Serif title over the top, caption pill below.



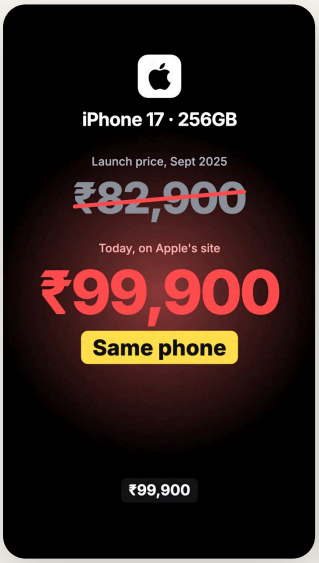
The big number. Counts up fast, with a yellow pill for context.



The real source. A page captured at phone width, highlighted, named and dated.



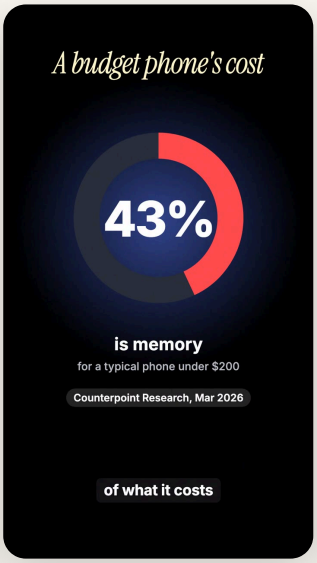
The product. Its real site in a phone frame, with logos of what it connects.



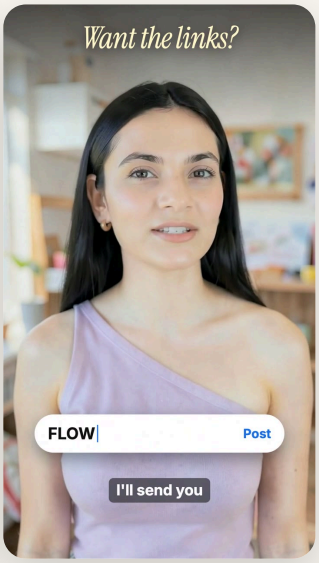
Before → after. Strike the old value, slam in the new one.



The chart. Two or three bars, sourced, with one hero colour.



The share. A ring that fills to the number she says.



The CTA. A comment box types the word as she says it.

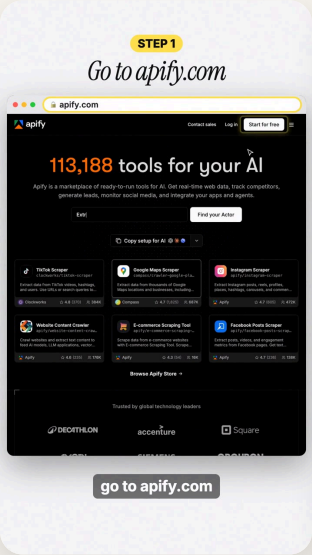
CAPTIONS

Small grey pills of up to three words, timed from a transcript of the final voice and placed low in the frame. Big yellow word-by-word captions looked cheap next to the reference creators.

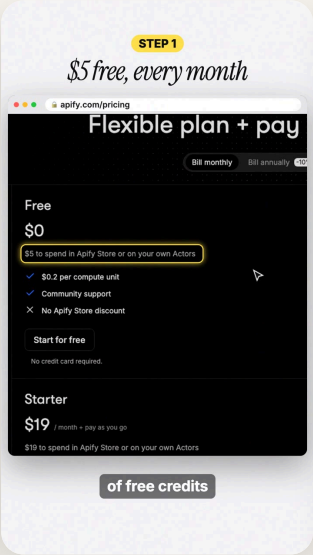
Tutorial steps people can follow

Our first tutorial failed because the screen recordings were zoomed onto the wrong things: viewers couldn't tell where they were. The fix is a strict pattern for every step:

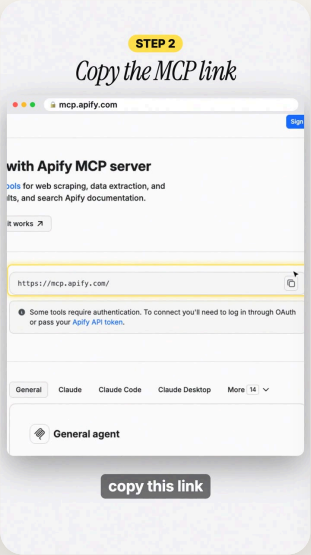
1. **Record at a tall window size** (960 × 1100) so pages stack neatly in a vertical frame.
2. **Show the address bar** with the real URL, and light it up when she says it.
3. **Show the whole page first** so viewers get their bearings, then zoom.
4. **Ring exactly what she names**, at the moment she names it. Place rings from the element's measured position, then check by eye.



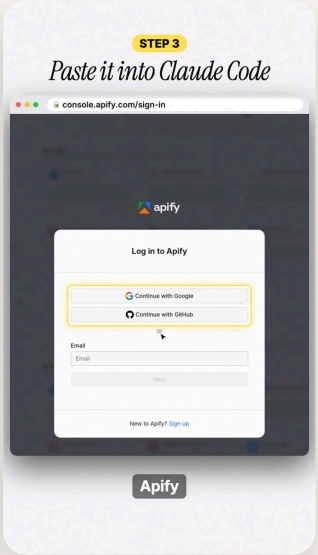
Step 1: the URL lights up on “apify dot com”.



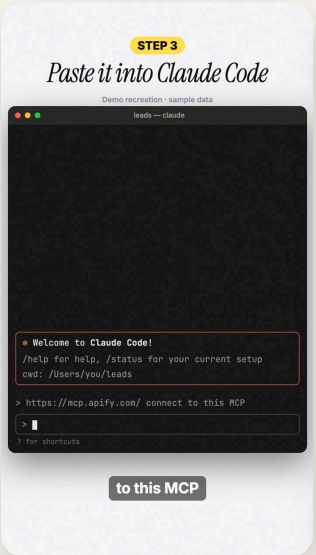
“$5 to spend” ringed on “five dollars”.



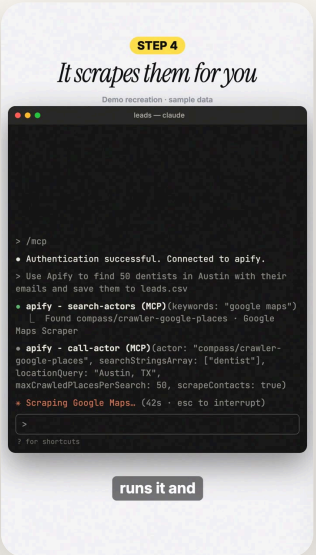
Step 2: the link box ringed, then copied.



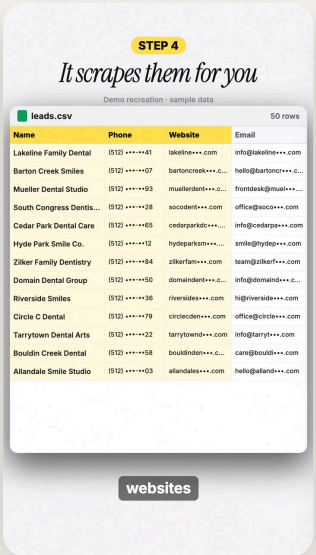
The real sign-in page, buttons ringed.



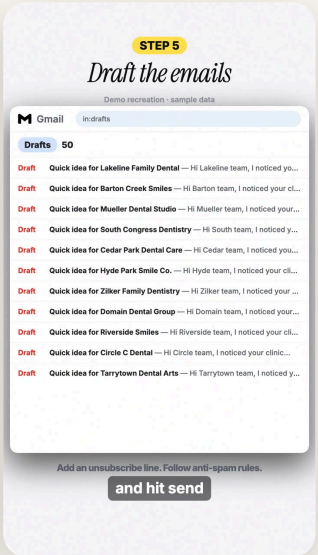
A terminal step, typed in time with her words.



Each tool call appears as she names it.



Columns light up: names, phones, websites, emails.



Drafts, never auto-sent, plus an anti-spam note.

Top row: real recordings of public pages. Bottom row: recreations of steps that need a login, tagged “demo recreation · sample data” on screen.

The last ten minutes matter most

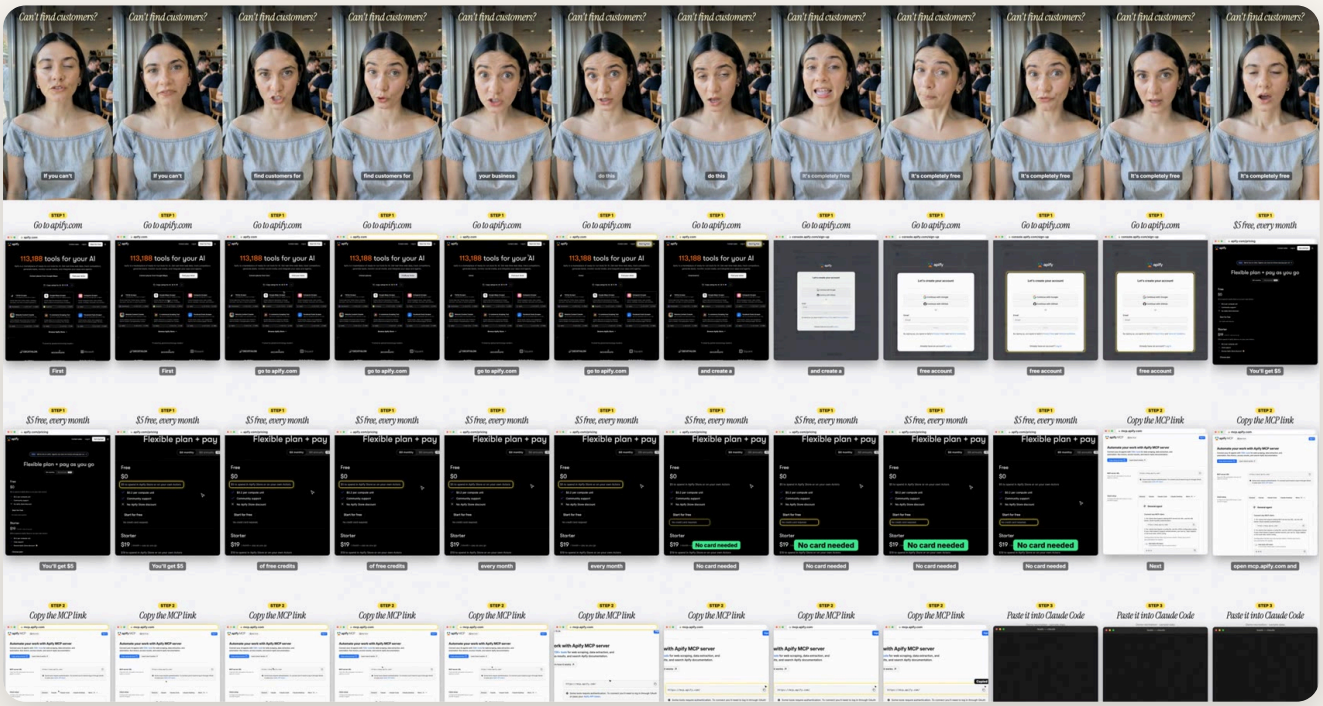
Mix

- Voice on top; music around 18–20 dB under it
- A short whoosh or tick on most cuts
- Free sound effects found online beat generated ones
- Final loudness around **−14 LUFS**, peaks below −1.5 dB

```
ffmpeg -i silent.mp4 -i voice.wav -i
music.wav \
 -filter_complex "[2]volume=0.12[m];[1]
[m]amix=2:normalize=0,
 loudnorm=I=-14:TP=-1.5[a]" -map 0:v -map "
[a]" \
 -c:v copy -c:a aac reel.mp4
```

Check (every single reel)

- [ ] Transcribe the finished reel and read it back. It catches misheard brand names and swapped words.
- [ ] Make a contact sheet (one frame every 0.4 s) and scan it for blinks, grimaces and graphics sitting too high.
- [ ] Watch the first 3 seconds on a phone: eyes open, lips moving, hook readable.
- [ ] Check every number against its source on screen.
- [ ] AI label on, sources in the caption, CTA word matches the end card.



A real contact sheet of a finished 53-second tutorial reel. One glance shows pacing, the on-camera share and any bad frames.

A rhythm that ships daily

The trick is to separate **thinking work** (scripts, picking, editing) from **machine work** (voices and talking shots), and let the machine work overnight.

WHEN	YOU	YOUR MAC
Evening	Research 3–5 topics, write scripts, pick outfits	Voice takes (~5 min per reel)
Night	Sleep	Talking shots: 2 takes × 3 lines × 5 reels (6–10 h)
Morning	Pick takes with the mouth strips (10 min per reel)	Idle
Day	Edit, mix, check (45 min per reel)	Renders the edits (3–5 min each)
Next evening	Schedule posts; reply to comments	Next batch

BATCH TIP

Queue every shot of every reel in one script, and set a completion alert for **each reel**, not just the first. We once had a finished reel sit unnoticed for two hours.

SPEED TIP

Keep the Mac plugged in and awake (`caffeinate -i`). When it throttles, a 15-minute shot takes 40 or more.

Bonus · Option B: use your own face

We also cloned a real face (with full consent: it was our own). What worked: start from a clean, well-lit photo of the real person rather than an AI redraw, which never looked like them. Then use the same audio-to-video recipe, with the calm prompt and lines kept to 2–3 s. What failed: lip-dubbing existing footage (the lips didn't match), and pasting a real face onto generated photos.

FIX-IT GUIDE

Every problem we hit, solved

YOU SEE	WHY IT HAPPENS	FIX
Lips barely move on the first words	Starting photo has an open grin	Use a calm, closed-lip still for every talking shot
First words look dubbed	Scene starts before the shot's audio does	Scene start = where the shot starts in the voice
Mouth stuck in an "O", or a frozen grin	Fast or smiling line; bad seed	Second seed; if both fail, use a title card
Grimace, clenched teeth, forehead lines	Prompt asks for big expressions	The calm BASE prompt in chapter 5
Eyes closed or squinting	Model blinks freely	"eyes wide enough open, blinking only rarely"; check frames
Head jumps mid-shot	Random motion in the take	Cut before the jump, or use the other take
A second phone in her hand	"selfie" or "phone" in the prompt	Remove those words; say where her hands are
Face or earrings drift	Shot too long	Lines under 5 s; join shots in the edit
Hands melt	Hands moving or tucked away	Out of frame, or holding a simple object
She looks older or oily	Hard light; "shine" in the prompt	Soft daylight; "the same youthful skin"
Voice rushed, no pauses	Cleanup removed all silences	Keep 0.3–0.5 s after each sentence
Brand name mispronounced	The voice model guesses spelling	Respell it ("Clawd") and regenerate that sentence
Wrong word in the voice	Take drifted from the script	Pick takes by transcript match
Screen steps unreadable	Desktop-size recording, zoomed blindly	Tall window, full page first, ring the target
Render 2–4× slower than usual	Mac throttling or swapping	Plug in, stay awake, one model at a time
Fake captions or logos in video	"no text, no logos" in the prompt	Don't mention text at all

PRINT THIS PAGE

The per-reel checklist

Script

- [ ] Hook lands in under 3 seconds
- [ ] Every number has a real, dated source
- [ ] No first-person claims or testimonials
- [ ] She's on camera for 2–3 lines only
- [ ] CTA: one word, and what they get

Voice

- [ ] Three takes; best transcript match picked
- [ ] Pause after every sentence
- [ ] Brand names sound right

Talking shots

- [ ] Calm, closed-lip still for this outfit
- [ ] Frame count is 8n + 1
- [ ] Two seeds per line

Picking

- [ ] Mouth strip checked: closes on m, b, p
- [ ] No grin, squint, lean or head jump
- [ ] Eyes open on the first frame

Edit

- [ ] Scene starts match shot starts
- [ ] Sources shown with name and date
- [ ] Tutorial steps: URL, full page, then ring
- [ ] Recreations tagged on screen
- [ ] Captions read correctly

Ship

- [ ] Loudness about −14 LUFS
- [ ] Transcript of the final reel read back
- [ ] Contact sheet scanned
- [ ] AI label on; sources in the caption



NOW BUILD YOURS

One face. One voice. A reel every day.

Start with chapter 2 tonight: generate eight faces and pick one. By the weekend you can have your first reel rendering while you sleep.

THE SOLO ENTREPRENEUR

youtube.com/@thesoloentrepreneur07

All people shown in this book are AI-generated. Tools and models belong to their respective owners and are used under their licences.