



Raw recording in. Finished edit out.

The exact prompt Opus 5.5 and GPT-6 Astra each got to edit the same talking-head reel. Same recording, same prompt, and no human edits to either result.

HOW TO USE IT

Paste the absolute path to your recording at the very end.

THE VIDEO

Swap the brand block for your colours and typeface, or delete it.

STEP 5

Send it once, then judge the step 4 table as closely as the render.

STEP 4

Needs Remotion, Whisper and FFmpeg. The model installs what's missing.

TOOLS

/goal Turn the raw recording at the path below into one finished, polished, upload ready video, edited entirely by you. Nobody is going to fix your output by hand, so the file you produce is the file that gets posted.

WHAT YOU ARE GETTING

A single talking head recording, already trimmed to its final spoken cut. No graphics, no captions, no motion, no titles. There is no script, no transcript, no shot list and no storyboard. You work out what it needs from the audio.

WHAT DONE LOOKS LIKE

One video file next to the source, named <source name>-edited.<ext>, that a person could post without touching it. Same resolution, frame rate, aspect ratio AND duration as the source, to the frame. Captions on screen and in sync. Motion graphics on the moments that earn them. If you would not post it yourself, it is not done.

STEP 1, LOOK AT WHAT YOU HAVE

Probe the source and record duration, resolution, frame rate and audio sample rate. Print them. Every later decision hangs off these numbers. If ffmpeg and ffprobe are not on this machine, install them.

STEP 2, TRANSCRIBE IT, LOCALLY, WITH WORD LEVEL TIMINGS

Use Whisper, running locally. Do not send the audio to a hosted API.

Work out what this machine already has before you install anything. Whisper comes in several forms and any of them is fine, so search for one that is already here, and only install one if there is nothing. Whichever you use, it has to emit per WORD timestamps, not just per sentence, and you may need a specific flag or output mode to get them.

Print which Whisper build, which model size and which flags you chose, and why, before you run it.

Then, whichever one you land on, these three things are true and they will silently ruin the edit if you skip them:

- VERIFY THE TRANSCRIPT COVERS THE WHOLE FILE. Compare the last timestamp against the real duration from step 1. A transcript that stops partway through parses perfectly and looks completely valid, and every cut you derive from it will be wrong.
- UNDERSTAND WHAT TIME BASE YOUR TIMINGS ARE IN. Tools that run voice activity detection often report offsets with the silence squeezed out, which do not map back onto the real file. Prove your word times line up with the source before you trust them, by spot checking a few words against the audio.
- TREAT WORD BOUNDARIES AS ESTIMATES. Speech to text stretches a word to fill the gap it lands in, so a word time is approximate. When a graphic has to hit a word exactly, check that word against the waveform rather than trusting the number.

STEP 3, DO NOT TOUCH THE TIMING

The cut is already made. The recording you have been given is the finished spoken cut, and it is locked.

- Do not remove a single frame. Do not tighten a pause, do not trim a stumble, do not shorten the top or the tail.
- Do not reorder anything, do not change the speed, and do not change the

pitch.

- The finished file is the exact same duration as the source, to the frame. That is a hard requirement and you should check it at the end.

You are editing the picture, not the timeline. Everything you add sits on top of a cut that does not move.

STEP 4, DECIDE WHAT THE VIEWER SEES

This step is the whole job and I am deliberately not telling you how to do it. Read the transcript as content, not as a caption source. Work out what this video is actually about, where it turns, and what a viewer needs to SEE at each moment to follow it. Then design that. The vocabulary of graphics is yours to choose. Four things I will say, because they are craft rather than taste:

- Anchor everything to the exact moment the words are spoken, using the word timings. A graphic that lands a beat late is worse than no graphic.
- Many moments need nothing. Restraint is part of this. A graphic on every line is worse than a graphic on the right lines, and knowing which lines to leave bare is most of the skill.
- Do not describe on screen what is already being said out loud. Show the thing, do not label it. A rectangle with a sentence in it is the failure mode, and it is what a lazy pass always produces.
- Design it so a viewer with the sound off still gets it.

Print your plan as a table of timecode, spoken line, and what you chose to show, with your reason. Include the moments you deliberately left bare, and why. I will judge the video on this table as much as on the render.

STEP 5, BUILD IT

Every graphic is built as code, in Remotion. Install it fresh from npm into your own working folder, the normal way. Read the API docs at the source repo when you need them:

<https://github.com/remotion-dev/remotion.git>

Do not clone it and do not build Remotion from source.

- Scaffold a Remotion project, pin the version, and match the composition to the source resolution and frame rate from step 1.
- The speaker's footage is the base layer for the whole video. It never disappears, and the speaker is never fully covered.
- Captions are required. Word level, in sync, legible on a phone, and never more than a handful of words on screen at once.
- Build a real component for each idea. Do not build one generic card and pass it 12 different strings.
- Nothing sits still for long. Everything enters and leaves deliberately, and nothing is left stranded on screen after its moment has passed.
- Respect a safe area. Keep text and graphics well clear of every edge, and clear of where a platform puts its own interface.
- Pick one typeface and one weight ladder and hold it for the whole video.

BRAND, REPLACE THIS BLOCK WITH YOUR OWN OR DELETE IT

Follow this exactly. If you delete it, choose a look that suits the content and hold it consistently.

Background: #f7f9fb, white #ffffff surfaces. Light, never dark.

Ink: #191c1e, dimmed #2a2838, muted #55535f

Hairlines: #c7c4d7, soft #e6e8ea
FILLS: burnt orange #c2410c, gradient:
linear-gradient(120deg,#c2410c,#d97757)
EMPHASIS TEXT: cyan blue #0284c7, gradient:
linear-gradient(120deg,#38bdf8,#0284c7,#075985)
Failure state: #D34135, sparingly and only for a real failure
Typeface: Inter, with a system mono stack for code and terminals
Shadows: barely there. 0 2px 14px rgba(25,28,30,.05)
Gradients are the default for a fill, not an accent. For gradient TEXT, clip the gradient to the glyphs and set color to transparent, or the background silently paints over the type.
Never use teal or green as a brand color. Never introduce a third accent.
One rule holds whether you keep this block or delete it: realism beats the brand palette for anything quoting a real interface. A terminal, a file tree, a code diff or an editor should look like the real thing, so terminal green stays green.

STEP 6, AUDIO

Leave the audio exactly as it is. Do not normalize it, do not compress it, do not add music and do not add sound effects. Pass the original audio track straight through to the output.

STEP 7, PROVE IT BEFORE YOU CALL IT DONE

A passing build is not evidence. Do all of these and report the results:

- Render actual still frames across the whole video and LOOK at them. Nothing overlapping, nothing off the safe area, nothing clipped, nothing blank.
- Sort every graphic's start and end time and diff adjacent pairs. Two graphics rendering over the same seconds is the most common defect in this kind of build, and it passes every test that does not explicitly compare them to each other.
- Confirm the output duration matches the source duration to the frame, and that the audio is byte for byte the audio you were given.
- Check the last caption's timestamp against the real duration. Captions stopping early means the transcript was truncated. Go back to step 2.
- Measure the final loudness and state the number.

Say plainly which of these you actually ran. Do not claim a result you did not observe.

WHAT I AM NOT GIVING YOU

No storyboard, no shot list, no direction on which moments deserve a graphic and no vocabulary of graphics to pick from. Those choices are the test. Make them, and defend each one in the table from step 4.

THE VIDEO:

<paste the absolute path to the recording here>