Claude als Video-Editor

Der Prompt, mit dem Claude aus deinem Rohvideo ein fertiges Reel schneidet.



Berkan Ceyhan · Stand Oktober 2026

Kopier den Prompt in Claude Code

Trag unter **INPUTS** dein Video und die Musik ein und schick den Prompt. Die Soundeffekte sucht Claude selbst. Danach arbeitet es ohne Rückfragen, bis `final.mp4` fertig ist.

NOIR & IVORY Edit

```
You're editing my raw talking-head recording (or an interview clip) into a finished, ready-to-post 9:16 short (Instagram Reel, TikTok, YouTube Short) in the NOIR & IVORY style: a premium, cinematic edit that alternates warm black and warm parchment designs, cuts and animates on the beat of the music I give you, and uses real sound effects that you find yourself. Do all of it yourself, end to end, in code: ffmpeg, Python, Node and HyperFrames (npx --yes hyperframes). No video editor app. You may install what you need locally inside this project (npm packages in the project folder, pip into a local venv, model weights inside the project folder), without asking. Never install system-wide software.
WORK FULLY AUTONOMOUSLY
Don't ask me anything after this message. Run every step start to finish and finish with final.mp4 exported.
Every judgement call (which take, where the music is cut, which line becomes a poster, which ground, which object stands for a word, which sample fits an event, a fact that disagrees with its sources) is yours: make the best call, keep going, and log it in the report.
If something fails, try another way (a different tool, method or fallback) rather than stopping. A missing tool is not a reason to stop: install it locally or use an alternative.
If the cover title field is empty, write one from the hook yourself.
Only stop early if the input files are missing or unreadable.
STYLE REFERENCE (study it before you plan, if you have one)
<optional: folder of a finished edit in this style, e.g. one you made earlier with this prompt> is a finished edit in this style: out/final.mp4, out/compare.mp4, out/REPORT.md (run table, beat snapping, SFX table, levels), graphics/lib/cin.css + cin.js (ground, type and motion system), graphics/*/scene.html (every composition), work/compose.py (transitions), work/audio.py (sample mapping). Match its look, rhythm and sound. Without a reference, the rules below are the complete style. Either way, build this video's graphics, music edit and mix yourself.
INPUTS
Camera video: <path to raw video, or a YouTube/Instagram URL plus the time range>
Clean audio: <path to separate mic WAV, or leave empty to use the video's audio>
Music: <path to the music file>
Cover title (frame 0): <your title, one or two lines, or leave empty>
Speaker credit: <e.g. "Jensen Huang · Lex Fridman Podcast #494">
```

```
Before/after version: <yes or no>
DELIVERABLES (in a new folder editNN_<topic>/ next to the other edits; never reuse an existing folder)
out/final.mp4: 1080x1920, 60 fps, H.264 (crf 16-18, -tune grain), AAC 48 kHz stereo 320 kb/s
out/stems/voice.wav, sfx.wav, riser.wav, music.wav (all full length, so I can remix)
out/thumbnail.png and out/thumbnail.jpg
out/compare.mp4 only if the before/after version is yes
out/REPORT.md, a short report in German (see the end)
SOURCE, SYNC AND TRANSCRIBE
If the video is a URL: download it with yt-dlp (uvx yt-dlp), video and audio separately at the best quality. An mp4 section download can carry an edit list (a pre-roll before t=0) that a stream copy into mkv silently drops and that shifts picture against sound by seconds: re-encode the video once while honouring the edit list (crf 8), and confirm the audio offset by cross-correlating against a combined download of the same section.
If a clean audio file is given, cross-correlate the camera audio and the clean audio at 16 kHz mono (confirm a clear peak) and use only the clean audio from here on.
Transcribe with word-level timestamps (faster-whisper large-v3; pass a numpy array, not a path). Refine every word boundary with MMS forced alignment (torchaudio) of the verbatim transcript; when whisper and the alignment disagree on a word, compare alignment scores of both readings and keep the better one.
THE CUT
I re-record lines until I'm happy, so keep the LAST complete take of every line; use an earlier take only if the last one has an audible defect. Remove restarts, self-corrections, repeated words ("every every", "I would, I would"), "okay", stray syllables and muttered comments. In an interview clip, also remove asides the short doesn't need, and log every removal.
Split every line into islands at quiet stretches over 250 ms (10 ms RMS under -38 dB, blips of up to 30 ms ignored); a join gets 120 ms of quiet where the source had more than 250 ms, the natural gap where it had less, 70 ms where words were removed inside running speech (cut at the quietest 10 ms window between the words).
Leave 40 ms before the first word. A line's end is its measured true end (last 10 ms window at or above -35 dB). Picture cuts are hard cuts on the island onsets, frame-quantised at 60 fps. The J-cut is in the audio only: each island's audio starts 50 ms before its picture cut; tails hold to the noise floor and fade over 170 ms, never past the next sound. No join may have a quiet stretch over 150 ms. Check the cut by transcribing the cut voice track: every kept word whole, nothing clipped.
END = the last frame of the voice track. Every layer ends exactly on END.
Footage: a fixed 9:16 crop per camera shot, centred on the speaker's face (YuNet), scaled to 1080x1920 with lanczos and a light luma unsharp if it is upscaled. No zooms, punch-ins or moves on the footage, ever.
EYE CONTACT AND CUTAWAYS
On my own recording: my footage shows only while I'm looking at the lens; a line read off to the side becomes a poster. On an interview clip: skip that rule, but detect every camera change; any shot that does not show the speaker (the host, a wide) must sit under a poster in the edit.
MUSIC: EDIT IT ON THE BEAT
Analyse the track first: tempo, beat grid, bars and downbeats (librosa, tightness 400-600), loudness curve, sections (intro, entries, the strongest downbeat, the ending) and key.
Edit it into pieces cut on downbeats (bar phase preserved, 10-120 ms equal-power overlaps) so that: the first strong entrance of the music lands on the hook-end join; the track's strongest downbeat lands on the join that opens the second half (the climax); the track's own ending lands on END (then a 1.5 s fade that ends on END). Start under the hook with the track's quiet intro (5 ms declick). Log every splice with track and video times.
ONE flat level, no ducking, no automation: -27 LUFS integrated (about 13 LU under the
```

```
voice). Stem: out/stems/music.wav.
Then detect the beat grid on the edited music stem itself (video time): beats, and
downbeats every 4 beats in phase with the hook-join entrance. Everything below that says
"on the beat" uses this grid.
THE LAYOUTS
Canvas 1080x1920. Three layouts:
a) Footage: the crop, full frame, with a soft vignette (edges -30 %) and fine luma grain
(±2.5/255), plus the footage overlays.
b) Poster: a full-frame design on one of two grounds, Onyx or Ivory.
c) Split: on Ivory, my footage inside a framed photograph (window x 90-990, y 230-1050, the
footage rows 160-1144 scaled to 900x820; a 14 px warm-white mat, radius 30, a 2 px gold
inner line, a soft brown shadow), with a list or a short statement building up below it on
the beat.
Rules: the hook is a poster on Onyx; footage and graphics alternate; a graphic run lasts 1-
4 bars, a footage run up to 4 bars; graphics take 50-65 % of the runtime; every run sits on
one ground; use the split at most once and for a list.
Grounds: alternate Onyx and Ivory so that about half the graphic runs are bright. Let
darkness and light carry meaning where the script allows it: the problem, the missing
skill, the old way in Onyx; the answer, the expert, the result in Ivory (a "light switch"
between them is the strongest moment). The climax may be either; the ending sits on
whichever ground lets its object shine.
Onyx: radial fill #241D16 at 50%/40% through #15110D (28 %) and #090706 (62 %) to #040303
at the edges; a warm light beam from the top (rgba(255,214,150,.13) falling off) and a soft
warm pool at mid height; 30-40 warm dust particles drifting up through the beam
(deterministic, pure functions of t). Text warm white #F2EADB, leads sand #A89C88, accent
gold #E2B661, kicker gold, credits #6E6455.
Ivory: radial fill #F4EDE0 to #E0D3BC; text ink #1C1712, leads #7D6F5C, accent dark gold
#9A7124.
Type: Cormorant Garamond for every keyword: titles in caps 600, 140-170 px, letter-spacing
.04 em; italics 500, 112-168 px, in gold for the benefit word. Inter only for small lines:
leads 500 40 px, kickers 600 30 px caps tracking .32 em, credits 500 30 px. Nothing under
30 px. Chapter marks at the top of each poster (Inter 600 30 px caps, gold hairlines either
side): Prologue, I · <name>, II · <name> ... Epilogue, naming the act, never restating the
words on screen.
Every spoken word of a graphic run appears on it in the order spoken; words already spoken
when a run starts show at once.
HERO OBJECTS
One dominant photoreal object per poster, rendered for the ground it sits on: on Onyx a
low-key studio shot on a dark charcoal backdrop with a warm golden rim light, on Ivory the
same object works too. Generate with an image model (Higgsfield MCP if it has credits;
otherwise run RealVisXL V5.0 locally through diffusers: single-file checkpoint downloaded
with 8 parallel curl ranges, enable_model_cpu_offload, vae tiling, 1024x1024, 32 steps, 4
seeds per object, pick the best from a contact sheet). If the model ignores a state
("unlit", "empty"), derive it from a render by image processing. Cut out with hyperframes
remove-background; where that fails (flat chips, gaps between handles, dark objects on dark
backdrops) use a saturation key or GrabCut, and check every cut-out on the ground colour
before using it. Log every prompt.
Every hero has, built in CSS from the same cut-out: a warm light pool on the back wall
behind it (radial rgba(255,205,140,.34) on Onyx, white on Ivory), a wall shadow (the
silhouette, brightness 0, blur 18 px, opacity .9 on Onyx and .30 on Ivory, offset +52/+74
px into that pool) that moves with 0.35-0.4x of the object's motion (parallax), and a
contact shadow ellipse that narrows and lightens as the object lifts.
Every hero floats: a sine of ±6-11 px over 2 bars, a slow sway of ±5 px, a rotation of ±1-2
degrees, and a small scale pulse on every downbeat. Vary the entrances and the special
motions across the video: rise from below, drop from the top, slide in from the left or the
```

```
right with a turn, a pendant lamp on a cord swinging over 2 bars, wheat or a plant swaying
around its base over 1 bar, a pen signing a gold signature stroke, a dead bulb that tries
to flicker on the beats.
MOTION ON THE BEAT
Snap every graphic join to the nearest beat in [-0.15 s, +0.30 s] around the speech-based
cut, preferring a downbeat in that window; log every snap with its offset. Title settles,
hero entrances and list items start on the nearest beat when their word is within 0.22 s of
it, otherwise on the word. On posters the camera steps instead of drifting: +1.2 % scale
and -3 px on every downbeat (0.45 s power2.out).
Words fade up legato (opacity, 10 px rise, 5 px blur, 0.38 s) at their spoken time. Titles
settle: letter-spacing closes from wide to its own value while they sharpen (0.5-0.9 s);
clamp the starting spacing to the free width so no word ever runs past 1000 px. Gold
hairlines draw out from their centre under titles. Exits blur out in 0.2-0.35 s.
FOOTAGE OVERLAYS
Plain lines: Inter 500 42 px warm white with a soft shadow, centred at y 1492-1530, in
chunks of 1-4 words that fade up word by word. The key phrase of a sentence: a kicker in
gold caps, a lead in Inter, the keyword in Cormorant (caps warm white, or italic gold) with
a gold hairline, all below the speaker's chin (measure it per run with YuNet, top edge at
least 40 px below it) and above y 1450, over a soft dark shade band. A list of three or
more items becomes either a rolling credits column (Cormorant 700 62 px caps, one fixed
slot per name, the column rolls up one slot 0.2 s before each new name, the previous name
at 42 %, the one before that fading out) or, on a split, a flowing sentence that builds up
item by item in Cormorant italic (new item in gold, turning ink after 0.5 s).
TRANSITIONS (compositor, 60 fps, J = the incoming run's first frame; all of them on beats)
Gold burn, for the big moments (the hook-end entrance, the climax, one or two more): J-
6..J-1 an amber film burn (#FF8F33 with #FFC46B added) grows from the top-right corner
while the frame gains +0.5 EV; J, J+1 a warm cream flash (#FFF1D6 at 90 %); J+2..J+9 the
incoming picture comes out of the warm overexposure while the corner burn fades; grain
throughout.
Dip through black, into a dark poster: J-5..J-1 fade to black, J..J+7 fade up.
Light switch, from a dark poster into a bright one: a hard cut on the beat, then J..J+7 the
incoming picture comes out of a soft white bloom (85 % falling off).
Warm light sweep, from a poster into footage: J-5..J-1 a warm glow gathers in the bottom-
left corner; J..J+7 a soft diagonal band of light (#FFD9A0) crosses the frame and the
incoming picture appears behind it while the last outgoing frame holds in front.
Hard cut, where the sentence runs straight on.
Use every type, no type twice in a row.
SFX: REAL SAMPLES YOU FIND YOURSELF
I don't give you a sound library. Find the samples yourself: download real recorded or
produced sound effects from free libraries whose license allows use in a commercial social
media video, with no attribution or with an attribution you put in the report (for example
CC0 sounds on Freesound, or sound effects under the Pixabay Content License). Check the
license of every file before you use it, keep the downloads in sfx/ inside the project
folder, and log source URL and license per file. No synthesis. Collect several candidates
for every role below, profile every file (length, active span, crest time, spectral
centroid, low-end share, pitch), look at spectrograms, and pick the best per role; leave
out what doesn't fit a premium, orchestral bed (glitches, shutters, mechanical counters,
cartoon pops). Roles and starting levels (peak relative to the voice's peak):
- intro riser (stem riser.wav, about -2 dB): a cinematic riser, pitch-shifted to the
music's key, cut at its crest so its loudest moment and its end land exactly on the hook-
end join;
- impact (-6): a deep cinematic boom on the hook-end entrance and on the climax downbeat;
- pre-riser (-7): the swell of a long cinematic whoosh, cut at its crest, ending on the
climax flash;
- burn swoosh (-5): airy swooshes, rotated, crest on the cream-flash frame;
```

```
- sweep whoosh (-10): a soft whoosh, crest when the light band crosses the centre (J+3.5);
- dip (-11 / -13): a swell ending in a low hit, or a low-passed reverb hit, on J;
- light switch (-17): a short bright sparkle on J;
- sub drop (-6): the deep boom on the punchline word that ends a list or a thought, two or
three per short, at least 4 s apart;
- gold keyword (-17 / -19): single bell strikes whose pitch fits the key, on italic gold
keywords;
- list items (-24): one isolated real key press each (cut a clean press out of a typing
recording);
- hero landings (-22): a soft pop, low-passed at 1.8 kHz, 0.4 s after the entrance starts.
Nothing on text reveals, hairlines, chapter marks or hard cuts. Measure every crest on the
high-passed envelope (not the raw peak) and place samples so that crest lands on its frame;
check it afterwards (all within ±10 ms). Then calibrate by loudness in the event's window
rather than by peak: an impact window about -17.5 LUFS and a burn window about -17 LUFS
(music included); keep sub drops 3-5 dB quieter where the music already carries heavy bass.
LEVELS AND MASTER
Voice (stems/voice.wav): -14.5 LUFS integrated measured on the stereo file, true peak at
most -3 dBFS (a transparent look-ahead limiter if the recording's crest factor is higher).
Music -27 LUFS. Master = the sum of the four stems, about -14 LUFS, true peak -1 dBTP at
most (a transparent limiter only on peaks). Measure every stem and the master with ffmpeg
ebur128 and report the numbers.
COLOR GRADE
My own camera footage: grade it to the same targets as graded footage (face 0.47-0.58,
black floor 0.008-0.020), warm and natural, measured before and after. Third-party footage
that is already graded (face 0.47-0.58, black floor 0.008-0.020): leave it ungraded and
report the measurements.
COVER
Frame 0 only: the hook poster in its finished state, built from the cover title (lead in
Inter, keyword in Cormorant italic, hero, gold caps line, credit). From frame 1 the hook
poster runs its own build.
BEFORE/AFTER (only if asked): out/compare.mp4, built in Python and ffmpeg: warm off-white
canvas #FBF8F2; "Rohes Video" (Cormorant Garamond SemiBold 72, ink) over the raw panel (x
100, y 688, 335x596, radius 18); "Claude Edit" on a gold #E2B661 marker over the edit (x
469, y 523, 506x900, radius 22); soft brown shadows; both frame-synchronous; audio =
final.mp4. The raw panel shows the kept takes on the same timeline, ungraded, without
overlays, and it must never show an empty frame: build a separate raw track in which every
cutaway shot (the host, a wide) is cropped around the person who is visible.
THUMBNAIL (thumbnail.png and thumbnail.jpg, quality 92+): 1080x1920 from the strongest
frame of the edit, with a short claim in the NOIR & IVORY type (cap height at least 135 px,
so it reads small in a grid) and the logos of any tools the video names; everything
important inside the centred 3:4 crop that Instagram shows in the profile grid.
QA BEFORE YOU FINISH (look at real frames)
A contact sheet of the whole video every 0.6 s and frame strips around every join: every
join on its beat, every transition reading as described, nothing clipped at the frame edge
during a title settle, no empty frame anywhere, every hero with a visible wall shadow,
contact shadow and float, the overlays clear of the face.
The beats: list each join's offset from its beat and each snapped element.
The audio: every SFX crest on its frame (±10 ms), the music's entrance on the hook join and
its strongest downbeat on the climax, no join gap over 150 ms, the levels above within 0.5
LU.
File size: if final.mp4 is over 150 MB, lower the grain or raise the crf.
Fix anything that fails and re-export. Don't hand it back until it passes.
REPORT (German)
END in ms; the source and how it was synced; take choices and every removal; the music
analysis and every splice; the beat grid and every snapped join; the runs table (layout,
```

```
ground, transition, chapter, keywords, hero and its motion); the source of every hero
(model, prompt, seed, cut-out method); the SFX table (time, file, role, level, event,
source URL, license) and the loudness calibration; each stem's and the master's LUFS and
true peak; the grade decision and numbers; the thumbnail's source frame, logos, claim and
the 135 px / 3:4 check; anything you couldn't do and what you did instead.
```

Du willst wissen, wo KI in deinem Betrieb Zeit spart? Buche einen Call mit mir und wir finden es gemeinsam raus.

Call buchen



Automatisierungs-Call für Agenturen, Coaches und Dienstleister

In 30 Minuten schauen wir gemeinsam, wo KI-Agenten und automatisierte Prozesse in deinem Betrieb zuerst Zeit und Geld sparen. Kostenlos.

FÜR WEN IST DAS?

- ✓ Agenturen, Coaches und Dienstleister
- ✓ ab 10.000 € Monatsumsatz
- ✓ ein klarer Flaschenhals im Tagesgeschäft
- ✓ du willst KI im Betrieb einsetzen

Call buchen

api.leadconnectorhq.com/widget/booking/mXsszCow3Sxqm4QEnOxK

Scannen & Termin wählen