

M A V E R I C K A I

CHATGPT CAN NOW EDIT VIDEOS FOR YOU

**CAPTIONS, SILENCE CUTS AND ON-SCREEN ANIMATION WITH
CODEX AND THE REMOTION PLUGIN**

Drag a raw talking-head clip into Codex, paste one prompt, get back a captioned, tightened, animated cut ready to post. Here's the setup, the prompt that actually works, what happens under the hood, and the honest first-run timeline.

Codex in the top left, Plugins, Remotion, Install: verified click by click

What your computer needs before the first edit (and what Remotion brings itself)

The video's prompt, expanded into one that gets it right first time

What Codex actually does: transcribe, find silences, cut, caption, animate, render

The approval gate that saves you from a bad render

Five more edits: the 30-second cut, the 16:9 version, B-roll text, jump-cut zooms, batch

Honest limits, render times, licensing, and the mistakes that stall people

2026 Edition

Your Editor Is Now A Prompt

Codex is the part of ChatGPT that writes and runs code on your computer. Remotion is a tool that makes videos out of code. Put them together and you get an editor that never opens a timeline: it transcribes your clip, finds every silence, cuts them, lays captions over the words as they're spoken, animates the moments that matter, and renders a finished MP4 into a folder on your Mac or PC.

I dragged a two-and-a-half-minute talking-head clip in, pasted the prompt in chapter 4, and got back a captioned, tightened, animated edit. Here's the part the video can't show in thirty seconds: what's happening while you wait, what your computer needs, and how to word the prompt so you don't get a robotic edit or a wall of questions.

1

prompt

0

timelines opened

30 min

first run, then about 10

HOW IT WORKS, IN ONE PARAGRAPH

Codex runs on your machine, so your video never leaves it. The Remotion plugin gives Codex a set of skills: how to start a Remotion project, how to caption, how to render. Codex writes a small program that describes your edit (which seconds to keep, where the captions go, when to zoom), then Remotion renders that program into a video using its own bundled copies of ffmpeg and a headless Chrome. Transcription runs locally with Whisper.

What it can and can't do

Does	Doesn't
Cut silences, ums and dead air	Generate new footage or B-roll
Word-by-word captions, styled any way you describe	Watch the video back; it reads numbers and frames, you are the QA
Zooms, pop-in headlines, progress bars, lower thirds	Add licensed music (you supply the file)
Re-crop and re-export for Reels, Shorts, YouTube	Turn vertical footage into real 16:9 (crop or pillarbox only)
Batch the same style across ten clips	Track a face for zooms unless you ask for it

Setup: Codex, Plugins, Remotion

Codex is included on every ChatGPT plan now, including Free, with usage limits that scale with your plan. The editing happens on your computer, so use the desktop app, not the browser.

STEP 1: Open Codex

In the ChatGPT desktop app, click the switcher in the **top left** and choose **Codex** (the third option after Chat and Work). Sign in with your ChatGPT account if it asks. Codex opens with a project view instead of a chat.

STEP 2: Install the Remotion plugin

Click **Plugins**, search **Remotion**, and click **Install plugin**. Or open [the Remotion plugin page](#) and click the install link there. There's nothing to configure. The plugin adds Remotion's skills to Codex: creating projects, captions, rendering, previewing in Studio, and its docs.

STEP 3: Check your computer once

Remotion brings its own ffmpeg and headless Chrome, so you don't install those. Two things it can't bring: **Node.js** (version 18 or newer, from nodejs.org) and, on a Mac, the **Xcode Command Line Tools**, which Whisper needs to build. Open Terminal and run **node -v**; if you get a version number you're set. If you're not sure about Xcode, tell Codex "check that node and the Xcode command line tools are installed and install what's missing" and approve the prompts. This is a one-time thing.

STEP 4: Make a project and drop the video in

In Codex, create a new project pointed at an empty folder (Desktop/edits is fine). Drag your video into the Codex window. Type **\$remotion** in the prompt and accept the plugin when it suggests it. Now paste the prompt.

HEADS UP

Codex asks permission before it runs commands on your machine. On the first edit expect a handful of "allow this?" prompts: cloning Whisper, building it, downloading a model, installing packages. Say yes to those. Say no to anything that isn't about this folder.

The First Run, Honestly

The video says thirty seconds. That's how long the prompt takes to type. The first edit on a fresh machine is twenty to forty-five minutes, mostly downloads and one render. Every edit after that is about ten, mostly render. Here's where the time goes so you don't think it's broken.

Step	First time	Every time after
Create the Remotion project, install packages	1 to 3 min	skipped
Build Whisper and download the speech model (medium.en is about 1.5 GB)	5 to 15 min	skipped
Transcribe a 2 to 3 minute clip (Apple Silicon)	30 to 90 s	30 to 90 s
Find silences, write the cut list, write the edit	5 to 15 min of Codex working and asking	3 to 5 min
Download Chrome headless shell (once)	1 to 2 min	skipped
Render 2 to 3 minutes of 1080x1920 at 30fps	3 to 6 min on Apple Silicon, 8 to 15 on Intel	same

What Codex is doing while you wait

It converts your audio to a format Whisper reads, transcribes it with word-level timestamps, runs a silence detector across the track and builds a list of the seconds to keep. Then it writes the edit as code: one segment per kept chunk, captions grouped three or four words at a time with the current word highlighted, and a zoom plus headline on each key moment. Then it renders. If it's good, it shows you the transcript with the cuts marked before rendering, which is what the prompt in the next chapter forces it to do.

TIP

Pick the Whisper model once. medium.en is the sweet spot for a single English speaker (about 95 to 98% word accuracy). small.en is a third the download and fine for clean audio. Names, brands and acronyms are where any model slips, and that's what the approval gate is for.

The Prompt That Works First Time

The video's one-liner works, but Codex comes back with four questions first: which platform, how long a silence, which transcription, what "animate" means. This version answers them and adds the approval gate. Swap the file name and brackets.

THE EDIT PROMPT

```
Edit [talking_head.mp4] for [Instagram Reels].

1) Transcribe it with word-level timestamps using
local whisper. If it isn't installed, download
whisper.cpp and the medium.en model without asking.

2) Cut every silence longer than 0.6 seconds and
every standalone "um" or "uh", leaving 120 ms of
room tone on each side of every cut so breaths
stay natural. No other audio processing.

3) Captions: 3 to 4 words per page, never a one-word
page, bold white sans-serif with a black outline,
current word highlighted in [#F5A623], block
centered at 70% of the height, max 900 px wide.

4) Pick the 5 strongest moments (a number, a claim,
a punchline), at least 10 s apart. On each: an 8%
zoom easing in over 0.4 s and back out, centered on
the speaker's face, plus a one-line headline under
24 characters that springs in at 22% height for
about 2.5 seconds.

5) Output 1080x1920 at 30 fps even if the source
isn't, H.264, AAC audio.

Before rendering, show me the transcript with every
cut marked and the 5 headlines as a plain table I
can edit. After I approve, render to
outputs/[name]_edited.mp4, tell me the final length,
and show me a still at each headline.
```

Why it's worded like that

120 ms of room tone is what "keep it natural" actually means; without a number you get a machine-gun edit. **Never a one-word page** stops orphan captions. **Centered on the speaker's face** matters because a phone-framed head sits high and a center zoom drifts it. **A still at each headline** is the QA step: Codex can't watch the video, only read it.

The Approval Gate

Before it renders, Codex hands you something like this. Real words and times come from your clip; this is the shape.

WHAT YOU GET BACK BEFORE THE RENDER

```
# TRANSCRIPT talking_head.mp4 (160.4 s)

00:00.0 Okay so [cut "um"] I want to talk about
00:03.1 something nobody tells you. [cut 0.9s]
00:04.9 I spent forty thousand dollars in my
00:07.4 first year and made back eleven. [cut 1.4s]
...

# CUTS: 41 silences (28.6 s) + 9 fillers (3.1 s)

# 160.4 s -> about 128.7 s

# 5 KEY MOMENTS

1 00:04.9 "$40K in. $11K back."
2 00:31.2 "Nobody says this part"
3 01:02.0 "26 months. One lesson."
4 01:41.6 "The 3-second rule"
5 02:19.0 "Do it before you're ready"

Reply approve, or edit any line.
```

This is where you fix the transcript (your brand name, a number it heard wrong), swap a headline, or un-cut a pause you wanted. Say "change headline 3 to 'This took me 2 years' and render" and it does exactly that. Nothing renders until you say so, so a bad transcript never becomes a bad video.

YOU ARE THE QA

Codex reads frames and numbers; it does not watch the result. Look at the stills it sends, then watch the MP4 once at full speed before posting. On a two-minute clip, expect to hand-fix two to five cuts that landed a hair early or late. Tell it the timestamps and it re-renders.

TIP

Ask for "npx remotion studio" and it opens a live preview in your browser where you can scrub the edit before rendering. Five seconds to open, and it catches caption overlaps for free.

Five More Edits

Same project, same chat. Each one is a follow-up, not a fresh setup.

THE 30-SECOND CUT

Make a 30-second version with only the best 3 moments, same caption and headline style. Include the sentence before and after each moment so nothing starts mid-thought. Hard cuts between them. Render to outputs/[name]_30s.mp4.

THE 16:9 VERSION

Make a 1920x1080 version of the full edit. Pillarbox the vertical video at full height with a blurred, darkened copy of itself filling the sides (no crop, no upscale). Captions at 82% height, max 1200 px wide. Headlines at 14%. Render to outputs/[name]_16x9.mp4.

JUMP-CUT ZOOMS

Alternate the framing at every cut between 100% and 106% zoom so the jump cuts read as intentional. Keep everything else the same and re-render.

PROGRESS BAR AND LOWER THIRD

Add a thin [#F5A623] progress bar along the bottom edge, and a lower third with [NAME] and [@HANDLE] that slides in at 2 seconds and out at 6. Re-render.

BATCH TEN CLIPS

Apply this exact edit style to every .mp4 in source/. Same rules, 5 moments each. One combined approval table, then render all to outputs/ and list the final lengths.

MUSIC

Drop a track in the folder and say how loud ("-18 dB under the voice, fade out over the last 2 seconds"). Codex can't source licensed music; for Reels, add the sound inside Instagram.

Limits, Licensing, And The Mistakes

Mistake	Fix
Using the browser version	Codex web runs in the cloud without your files. Use the desktop app for local edits.
"Cut every silence" with no number	0.6 s threshold plus 120 ms room tone. Otherwise it's robotic or it asks.
No approval gate	Add "show me the transcript and headlines before rendering." A wrong brand name in captions is a re-render you didn't need.
Declining the setup prompts	Whisper build, model download and Chrome shell are all one-time. Say yes to those.
Expecting 16:9 from a vertical clip	There is no hidden footage. Pillarbox with blur or center crop; say which.
Trusting captions on names and numbers	Whisper mishears brands and turns "forty thousand" into words. Fix them in the gate.
Rendering at max concurrency	More than about 6 on a laptop gets slower, not faster. Let Codex pick.
Posting without watching	Watch it once. You're the only one who will.

HEADS UP

Licensing. Remotion is free for individuals and small teams (companies up to three people). Bigger companies need Remotion's company license; check remotion.dev/license before this ends up in an agency workflow. Fonts from Google Fonts are fine for commercial use.

The one-page version

Step	Where	What
1	ChatGPT desktop app	Top left switcher, Codex
2	Codex, Plugins	Search Remotion, Install plugin
3	Terminal, once	node -v (install Node 18+ if missing); Mac: Xcode command line tools
4	Codex	New project on an empty folder, drag the video in, type \$remotion
5	Same chat	Paste the edit prompt. Approve the setup prompts. Wait for the transcript table.
6	Same chat	Fix names and headlines, say render. Check the stills. Watch the MP4. Post.

DO THIS TODAY

Install the plugin, drop in the last clip you filmed, and paste the edit prompt. Go make coffee during the first run. Every edit after that is a ten-minute prompt.

Want More AI Tips Like This?

I post daily AI content, tool breakdowns, and prompts that actually help you get more done. Follow me for the latest.

@MavGPT

TikTok | Instagram | Facebook

Film it. Drop it in. Approve the cuts. Post.