

K B V P U B L I S H I N G

THE FULL SYSTEM

The Viral Street Interview Ad Kit

How to build a 30-second AI street-interview UGC ad that doesn't look AI — from the flip-axis idea to a finished, checked clip.

\$27 · ONE-TIME PURCHASE

KBVPUBLISHING.COM



Welcome

This kit is one continuous system, in that order: the formula, the one idea that makes it work, how to write dialogue that doesn't sound written, the full production workflow, the prompt library, a real worked example end to end, how to post it or run it as a paid ad, and a place to get help if you get stuck. Work through it in order the first time — after that, it's a reference you'll come back to for every new product.

What's inside

MODULE 1

The Street Interview Formula

The five-beat shape every version of this ad follows, and why it works.

MODULE 2

The Flip Axis

The one idea that makes or breaks the ad — turning a benefit into the failure it solves.

MODULE 3

Writing Dialogue That Doesn't Sound Written

The rules that decide whether the finished ad reads as real or as a script.

MODULE 4

The Production Workflow

Product photo to finished clip, step by step, using the Master Prompt.

MODULE 5

The Prompt Library

The realism template and the five structural rules that stop a generation reading as AI.

MODULE 6

A Real Worked Example

The actual Halden production, start to finish — prompt, dialogue, stills, and the checks run before it was considered usable.

MODULE 7

Posting & Ad Strategy

Running this format organically versus as paid creative on TikTok and Meta.

MODULE 8

Community & Learning Hub

Where to get feedback on a generation, and see this format applied to other products.

APPENDIX

The Full Reference Prompt

The complete prompt structure behind the Master Prompt, ready to copy per product.

How to use this kit. If you just want a finished clip today, skip to Module 4 and follow the numbered walkthrough — it references Modules 2, 3 and 5 exactly where you need them. If you want to understand why the format works before you build anything, read Modules 1–3 first.

1. The Street Interview Formula

The "stop a stranger on the street" ad is one of the highest-converting UGC formats running on TikTok and Meta right now — and one of the easiest to fake convincingly with AI video, because the whole appeal is *rawness*, not polish. A slightly shaky handheld shot, a stranger who clearly wasn't expecting to be on camera, and a product reveal that surprises them. That imperfection is what makes people trust it.

Every version of this ad follows the same five-beat shape, in exactly 30 seconds:

1. **Hook (0–3.5s)** — a yes/no question the stranger is about to lose
2. **Pain (3.5–13s)** — they admit it, then tell a short, rambling, true story about the thing that doesn't work
3. **Sell (13–18s)** — the product is revealed and named, with one real benefit
4. **Reaction (18–24.5s)** — they try it or hear the claim, react, and commit — in one line
5. **Close (24.5–30s)** — the product is named again, where to buy it, and they walk off saying they'll buy it

That's the whole ad. Nothing more fits in 30 seconds, and trying to cram in extra beats is the single biggest reason AI-generated versions of this format fall flat — the beats get rushed, the subject stops sounding like a real person, and it reads as a script instead of a conversation.

Why this, and why now

This isn't a niche tactic. UGC-style ads now convert roughly 4x higher on click-through and cost about half as much per click as standard brand content, and street-interview-style creative specifically has become the default for performance ads in 2026 — native, unpolished formats are consistently beating polished brand ads on retention and conversion across TikTok and Meta. On the production side, the tools have caught up: AI-generated video is projected to make up over 40% of all video ad creative by 2027, and roughly 9 in 10 advertisers are now using or piloting generative AI for video. The window where this is still a genuine edge — rather than something every account is already doing — is now, not later.

One honest caveat. A portion of viewers say they're less likely to buy from a brand once they clock that an ad is AI-made. That's exactly why every rule in this kit exists — the whole point of the realism system below is that it shouldn't read as AI at all. Get the details right and nobody watching is thinking about how it was made.

2. The Flip Axis

Every good street interview ad hangs on one specific complaint getting flipped on one axis. Not three benefits. One.

The formula: take your product's main benefit, and invert it into the failure it's solving.

Your product's strongest feature usually gets written down as a benefit — "fast energy on the go," "keeps hair clean for 3 days," "saves you time." But nobody complains about a benefit. To write a pain beat that feels real, flip that benefit into the specific way its *absence* shows up in someone's life:

Benefit as written

The failure it implies

"Fast energy on the go"	You're out, you're flagging, and there's nothing on you
"Keeps hair clean 3 full days"	It's greasy by day two and you're washing it every morning
"Saves you time in the morning"	You're doing it every day and it eats twenty minutes

Once you have that failure, the whole ad writes itself around it: the **hook question** sets up that exact failure ("Be honest — do you crash in the afternoon?"), the **pain beat** is a stranger admitting to it unprompted, the **reaction beat** flips that exact axis back and nothing else, and the **close** restates the flip as resolved.

Golden rule. The stranger never sounds like they already know the brand. They're complaining about a *category* — energy on the go, greasy hair, mornings — and the product happens to solve it. The moment they sound informed about the brand, the illusion breaks.

3. Writing Dialogue That Doesn't Sound Written

This is the module that decides whether your ad works. Everything else — the camera, the lighting, the prompt engineering — is scaffolding around 95 words of dialogue.

Say the true thing plainly. Never write a joke.

The best line in this entire ad format is five words: *"It tastes like a lawn."* Followed by: *"Like actual grass."* That's not clever copywriting — it's a flatly stated true thing, repeated once in plainer words. That repetition is what a real, unscripted person sounds like.

The instant you reach for a clever metaphor ("you could fry an egg on it"), it stops sounding like a person and starts sounding like an ad script. The test isn't "is this a good line" — it's **"would this person actually say this to a friend?"** If it sounds written, cut it.

The AI's first draft will oversell — budget ten minutes to fix it

If you use AI to help draft your script, expect its first pass to sound like copywriting, not conversation. Ask it to describe a fragrance and it'll hand you something like *"bergamot and black pepper over cedar"* — technically accurate, completely wrong for this format. Nobody describing a scent to a friend on the street talks like a fragrance ad. They say *"yeah, it's nice, bit sweet."*

Budget ten minutes after any AI-assisted draft to read every line out loud and ask "would a real person actually say this?" — then rewrite anything that sounds written. That one pass, done by you, is often the single biggest difference between a clip that reads as an ad and one that reads as real.

Every line has exactly one job

You have roughly 95 words for the whole 30 seconds. Every single line must do one of four jobs:

1. **Hook** — makes them stop scrolling
2. **Pain** — makes the viewer recognise themselves
3. **Sell** — names the product and what it does
4. **Act** — tells them where to get it

If a line does none of those four things — a joke, a tangent about their day, a reaction that reacts to nothing specific — cut it. No exceptions.

Combine the reaction, the verdict, and the payoff into one line

Thirty seconds doesn't have room for three separate beats here. One line has to do all three jobs at once:

"Oh, that's actually nice. Yeah, I'd keep a pack in my bag — that's my three o'clock sorted."

Reaction ("that's actually nice"), commitment ("I'd keep a pack in my bag"), and payoff answering the original hook question ("my three o'clock sorted") — one breath, no filler.

The close is the second most important line in the video

A bare "Amazon." is not a close. The last three seconds are what the viewer walks away remembering, so the close needs to: say the product name again (minimum twice total), say where to buy it with real spoken padding — *"they're on Amazon, and there's a discount on right now"*, never a robotic "Amazon. 20% off." — and end on the stranger's purchase intent as the final line. Never end on a pleasantry like "thanks" or "cheers" — that closes the conversation, not the sale.

Never compress a line to hit a word count

This is the single most common way people break this format. When you're over the word budget, the instinct is to strip filler words — "yeah," "honestly," "like," "and." Don't. Those filler words are exactly what makes a line sound spoken instead of read. **When you're over budget, delete a whole line — never squeeze the words out of a line you're keeping.**

4. The Production Workflow

This workflow is tool-agnostic — it works with Seedance 2.5, or any other AI video model that accepts a reference image and a detailed text prompt. We built and tested this kit using Claude and Kie.ai specifically, and that's what the walkthrough below shows — but the prompt itself is just text. Any AI chat tool that can hold a long, detailed instruction and answer a few questions back (ChatGPT, Gemini, whatever you already pay for) fills it in the same way.

Step 1 — Get one clean product photo

A plain, well-lit photo of your product on a neutral background is all you need. This becomes the reference image the AI model locks onto for the product reveal. **Use your own product, or an invented one — never a real branded product you don't own or have permission to use.** An AI-generated stranger appearing to endorse a real commercial product he was never given is a fabricated endorsement, on top of the trademark risk of putting someone else's brand in paid media. Module 6 covers this in full.



The actual reference photo behind the Halden example running through this kit. Even lighting, plain background, one product, centred — nothing more is needed.

Step 2 — Pick your flip axis and write your script

Use Module 2 to find your flip axis and Module 3's rules to write your 95-word script.

Step 3 — Generate your prompt with the Master Prompt

Instead of filling in the realism template by hand, paste the Master Prompt below into Claude (or ChatGPT, Gemini, or any AI chat tool) in a fresh conversation. It asks six short questions about your product, then hands back a complete, ready-to-paste Seedance prompt plus a timed dialogue script.

The Master Prompt — copy everything below and paste it into a new chat

You are helping me write a prompt for an AI video model (Seedance 2.5 or similar) to generate a 30-second, one-take "street interview" style UGC ad. Ask me exactly these six questions, one message, nothing else:

1. Product name
2. Product description (key benefits/features - optional)
3. Main product benefit (what makes it different)
4. Main pain point it solves
5. Age / gender / ethnicity of the person being interviewed
6. Where can people buy it, and any current discount

Once I answer, do not ask any follow-up questions - decide the street, weather, wardrobe, and every other detail yourself, silently, and go straight to the finished output.

RULES FOR THE SCRIPT: invert the pain point into the specific way its absence shows up in someone's life; write plain, true statements, never a clever metaphor or joke; every line does exactly one job (hook, pain, sell, or act); combine the reaction, verdict, and purchase-intent payoff into ONE line; the close must say the product name again, say where to buy it with real spoken padding, and end on purchase intent, never a pleasantry; target 85-95 words total, and if over budget delete a whole line rather than stripping filler words; the subject hands the product back on the closing line and keeps walking past the camera, never turning back; direction stays restrained - one physical action and one micro-expression per beat; never describe the product's label or packaging in words, refer to it only as "the bottle" / "the packet" / "the tin".

BEAT TIMING (30 seconds, one continuous take, no cuts): 0.0-3.5s Hook | 3.5-13s Pain | 13-18s Sell | 18-24.5s Reaction+verdict+payoff | 24.5-30s Close

Then fill in every [BRACKET] in the reference template I'm about to paste below and deliver it as one copyable block, followed by the dialogue as a table (time / speaker / line / word count):

[PASTE THE FULL REFERENCE PROMPT FROM THE APPENDIX HERE]

The Appendix at the end of this kit has the full reference template referenced in the last line above — paste that whole block in underneath the instructions when you use this. It only has to be assembled once; save it as a note or template on your end and it's ready to reuse for every future product.

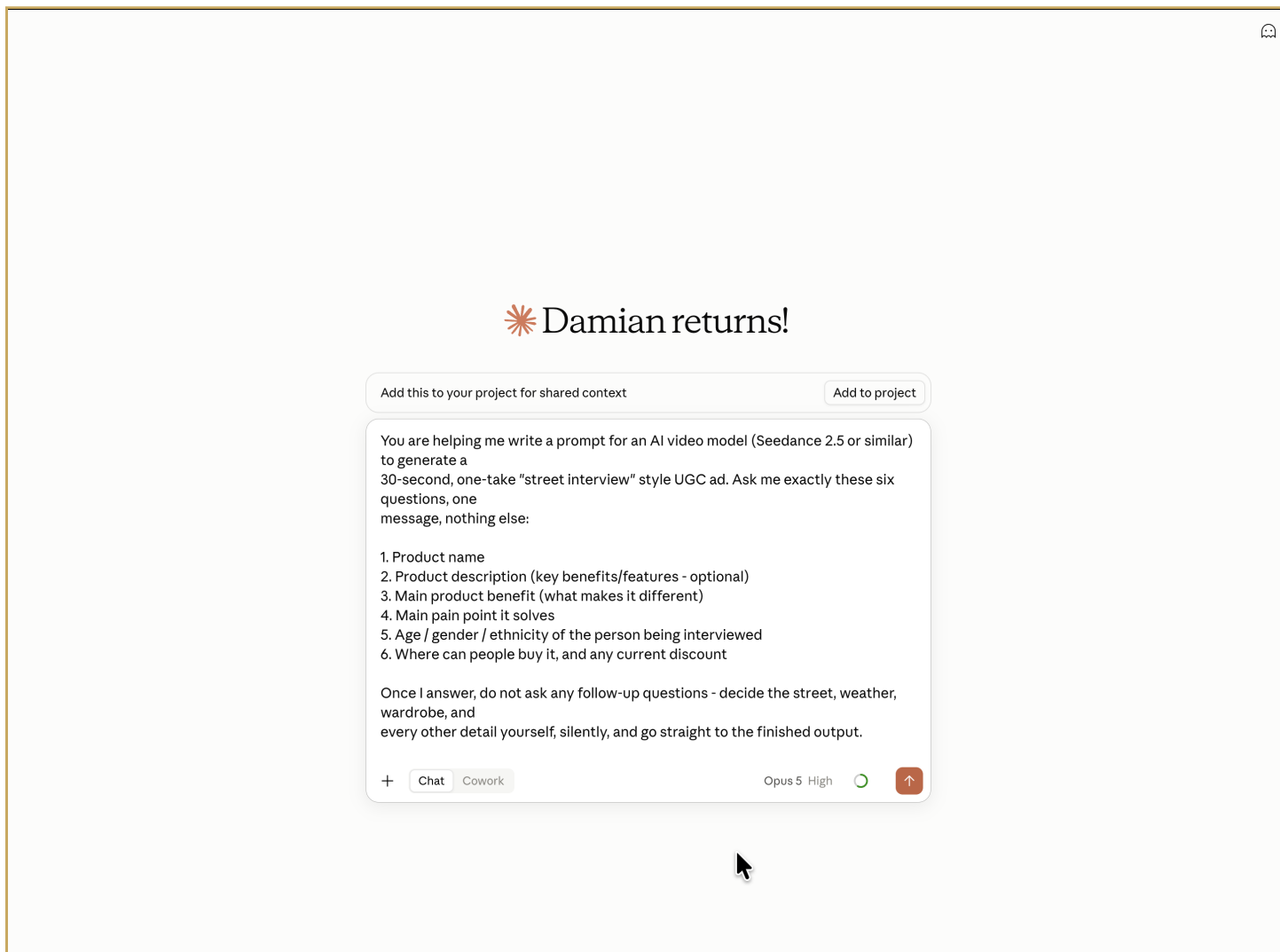
Here's what that looks like, step by step (using the Halden example from Module 6)

1. **Open a fresh chat with Claude** (or whichever AI tool you're using), paste in the Master Prompt, and send it.
2. **Claude comes back with the six intake questions** — nothing else, no side conversation.
3. **Answer with your product's real details.** For Halden: Product name — Halden. Main benefit — bergamot and black pepper over cedar, still going strong at the end of the day. Main pain point — most aftershave has faded by early afternoon. Age/gender/ethnicity — whoever fits your target audience. Where to buy — Amazon.
4. **Claude hands back the finished prompt and dialogue script in one message** — location, camera direction, subject and interviewer description, the anti-glitch and product-continuity rules, and the script itself, laid out with timings.

5. **Read the dialogue out loud before you do anything else with it.** If a line sounds written instead of spoken, don't rewrite it yourself — tell Claude what's off ("line three's too stiff, make it sound like he's actually moaning about it to a mate") and let it revise the line in the same chat. Because there's no separate skill file to upload, every fix is just a normal back-and-forth message.
6. **Do the same check on the location.** If what comes back doesn't fit the brand, say so and ask Claude to swap it. It rewrites that section without disturbing anything else in the prompt.
7. **Copy the finished prompt out of the chat** once the dialogue and setting both read right.

What this actually looks like in a real chat

SCREENSHOT 1



The Master Prompt pasted into a fresh Claude chat, ready to send.



FINAL TEST: If it feels like an ad, fail. If it feels like polished UGC, fail. If it feels like someone reading a script, fail. If he sounds like he already knows the brand, fail. If the product changes or warps in any frame, fail. If the brand name is mispronounced or differs between its two mentions, fail. If anyone walks past in the background, fail. If he walks off still holding the product, fail. If it feels like a stranger stopped on their lunch break who is quietly annoyed that their own version of this never works — pass.

Show less

just now   

1. Product name?
2. Product description (key benefits/features) — optional?
3. Main product benefit (what makes it different)?
4. Main pain point it solves?
5. Age / gender / ethnicity of the person being interviewed?
6. Where can people buy it, and any current discount?



+ Write a message...



Claude is AI and can make mistakes. Please double-check responses.

Sonnet 5 High

Claude comes straight back with the six intake questions and nothing else.

You are helping me write a prompt for an AI video model (Seedance 2.5 or similar) to generate a 30-second, one-take "street interview" style UGC ad. Ask me exactly these six questions, one message, nothing else:

1. Product name
2. Product description (key benefits/features - optional)
3. Main product benefit (what makes it different)
4. Main pain point it solves

Show more

1. Product name?
2. Product description (key benefits/features) — optional?
3. Main product benefit (what makes it different)?
4. Main pain point it solves?
5. Age / gender / ethnicity of the person being interviewed?
6. Where can people buy it, and any current discount?

🗉 🔍 🗑️ 🔄



- Halden
- Eau de toilette, bergamot and black pepper over cedar
- Bergamot and black pepper over cedar, still going strong at the end of the day
- Most aftershave has faded by early afternoon
- Male, late 20s, Black British
- Amazon, no current discount

+

↩

Claude is AI and can make mistakes. Please double-check responses.

Sonnet 5 High



The six answers for Halden, typed straight into the reply.

SEEDANCE 2.5 – STREET INTERVIEW, 9:16 VERTICAL, 30 SECONDS, ONE CONTINUOUS TAKE, NO CUTS

CAPTURE: iPhone 15 Pro rear camera, handheld by a person standing ~1.5m from the subject. Chest height, angled very slightly up, a degree or two off level. Constant handheld micro-shake from breathing and weight shifts. Natural drift and re-centring. No stabilisation, no zooms, no reframing, no cuts.

LOCATION: A wide UK high street, early afternoon. Overcast, flat diffused daylight from directly above, no visible sun, no hard shadows. A scratched perspex bus shelter, a shuttered betting shop with graffiti tags on the metal shutter, a delivery bike locked to a lamppost with its pannier still attached, a bin with a flattened cardboard box wedged on top, and faded black chewing-gum spots pitting the paving stones. Every vehicle is a single flat colour with completely blank panels – no livery, no lettering, no logo, no stripe.

Claude's reply — the finished SEEDANCE 2.5 prompt, filled in end to end.

Edit dialogue

"Excuse me — quick one. Can you actually smell your aftershave right now?"

to

"Hey There— quick question, Can you still smell your aftershave rthat you put on this morning"

Edit Location from A wide UK high street, early afternoon. Overcast, flat diffused daylight from directly

above, no visible sun, no hard shadows. A scratched perspex bus shelter, a shuttered betting shop

to

A Traditional US Shopping centre, early morning , Bright Interior lighting, with flat diffused daylight from directly above, no visible sun, coming through the roof no hard shadows. A selection of US shops however not text or logos or identifiable branding in shot

ⓘ Cancel Save

Fixing a typo while spotting an incomplete location edit. >

Updated the hook line — here's the dialogue table with that change:

Time	Speaker	Line	Words
0.0–3.5s	Interviewer	"Hey there — quick question, can you still smell your aftershave that you put on this morning?"	16
3.5–13.0s	Subject	"Nah, gone. It's always gone by about one. I end up carrying a bottle in my bag, topping it up all day."	22
13.0–18.0s	Interviewer	"Have you tried Halden? It's bergamot and black pepper over cedar, and it's still going by the end of the day."	21
18.0–24.5s	Subject	"Oh, that's proper spice, that. If it's still there tonight, I'm never carrying a top-up bottle again."	17
24.5–27.5s	Interviewer	"It's called Halden — you can find it on Amazon."	9

Asking for a line and a location revision in plain language, and the updated dialogue table that comes back.

SEEDANCE 2.5 – STREET INTERVIEW, 9:16 VERTICAL, 30 SECONDS, ONE CONTINUOUS TAKE, CUTS



Copy

CAPTURE: iPhone 15 Pro rear camera, handheld by a person standing ~1.5m from the subject. Chest height, angled very slightly up, a degree or two off level. Constant handheld micro-shake from breathing and weight shifts. Natural drift and re-centring. No stabilisation, no zooms, no reframing, no cuts.

LOCATION: A wide UK high street, early afternoon. Overcast, flat diffused daylight from directly above, no visible sun, no hard shadows. A scratched perspex bus shelter, a shuttered betting shop with graffiti tags on the metal shutter, a delivery bike locked to a lamppost with its pannier still attached, a bin with a flattened cardboard box wedged on top, and faded black chewing-gum spots pitting the paving stones. Every vehicle is a single flat colour with completely blank panels – no livery, no lettering, no logo, no stripe.

NOBODY WALKS PAST DURING THE TAKE. Background figures are stationary and distant only. No readable text and no brand identity anywhere in the background except the product itself.

FRAMING ARC (one take, operator movement only): The subject walks toward camera along the pavement, camera holding wide as he approaches. At roughly 2s he registers being addressed and begins to slow, settling to a stop at chest-height medium framing by 4s – this holds, with natural handheld drift, through the hook and pain beats. At 13s the interviewer's other hand enters from screen-left

The finished prompt, ready to copy straight out of the chat.

Step 4 — Generate it in Kie.ai

This walkthrough uses Seedance 2.5, because it's currently one of the few models that will generate a full, continuous 30-second take from a single prompt. You can use a different model if you prefer — the prompt structure carries over — but check its maximum single-generation length first. If a model caps out shorter, you'll need to generate the ad in multiple segments and stitch them together afterward in CapCut or a similar editor to reach the full 30 seconds. That stitching process is outside the scope of this guide.

1. Go to kie.ai, click **Explore AI APIs**, and search for **Seedance 2.5** (not 2.0 — 2.5 supports a full 30-second generation in one shot).
2. Paste your finished prompt into the prompt box, then remove any placeholder image, video, or audio files sitting in the form by default.
3. Set the resolution to **720p**, the aspect ratio to **9:16** (vertical), and the duration to **30 seconds**.
4. **Before you hit Run — upload your product reference image** from Step 1. This is the step it's easiest to forget, and without it the model has nothing to lock the product's look onto.
5. Hit **Run**. Generation typically takes 5–10 minutes.

6. **Review the finished clip in full before you use it anywhere.** Check: does the product stay the same size and shape throughout, do both hands keep five fingers, does anyone walk through the background (the #1 cause of warped faces), does the subject's face stay consistent start to finish, and does the subject walk past the camera on the close rather than turning back. Module 6 shows the full checklist run against the finished Halden clip.

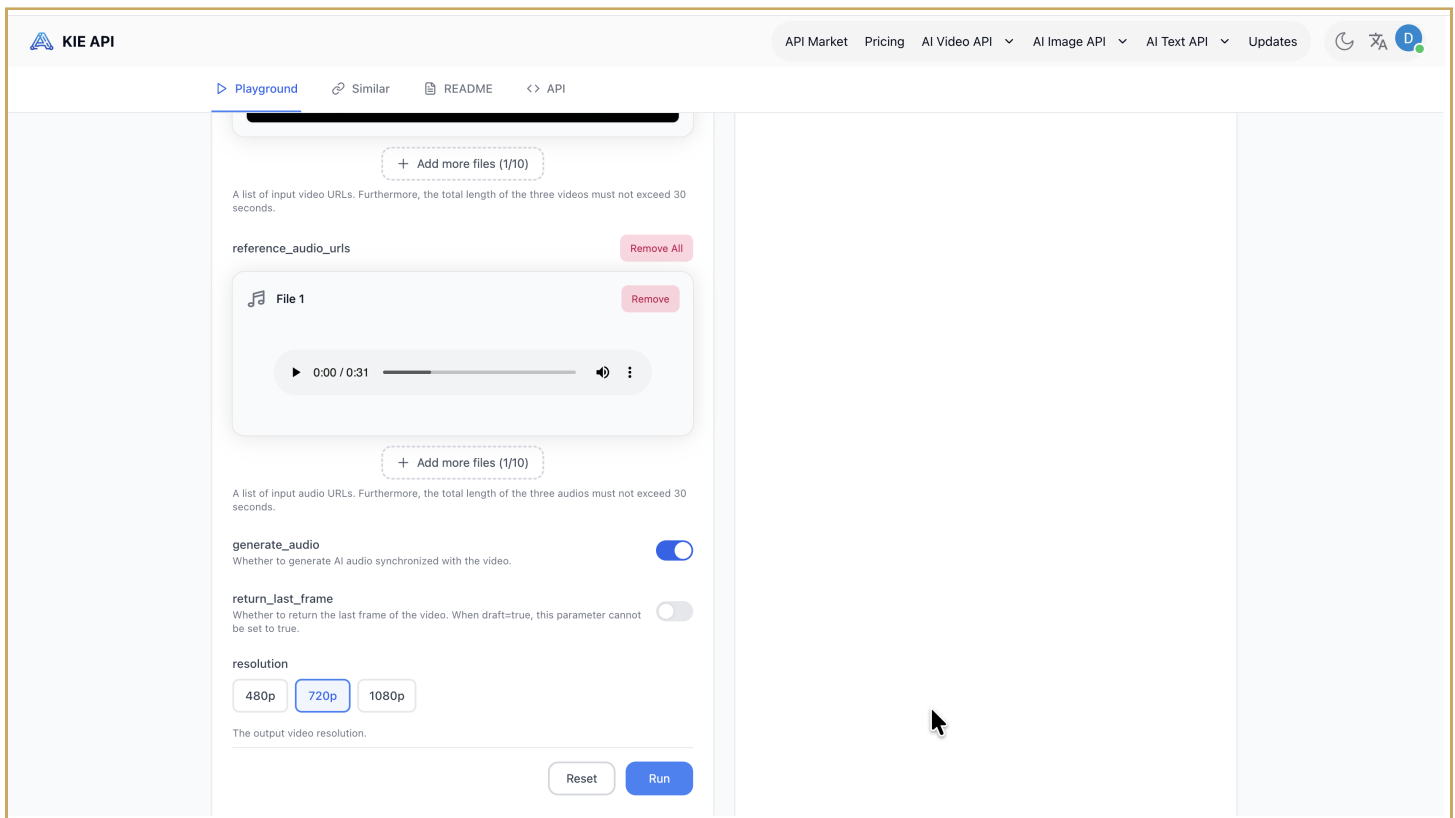
Pro tip — skip the manual re-sync. If you already know which voice you want before you generate, you don't have to swap it in after the fact with ElevenLabs Voice Changer (Step 4b) and re-sync it yourself in CapCut. Kie.ai's Seedance form has a **reference_audio_urls** field — upload a finished voice track there (for example, a line read out in ElevenLabs Text to Speech using your chosen voice) with **generate_audio** switched on, and Kie.ai generates the clip already lip-synced to that audio. This only works if you have the voice track ready before you generate; if you're swapping the voice on an already- finished clip, Step 4b below is still the way to do it.

SCREENSHOT 18

The screenshot shows the Kie API web interface. At the top, there's a navigation bar with links for API Market, Pricing, AI Video API, AI Image API, AI Text API, and Updates. Below this is a sub-header with 'Playground', 'Similar', 'README', and 'API' links. The main form area is titled 'reference_audio_urls' and contains a file upload section with a '+ Add more files (1/10)' button. Below the upload section, there's a toggle for 'generate_audio' which is currently turned on. There's also a 'return_last_frame' toggle which is turned off. At the bottom, there's a 'resolution' section with buttons for 480p, 720p (which is selected), and 1080p. At the very bottom of the form are 'Reset' and 'Run' buttons.

The reference_audio_urls upload field on Kie.ai's Seedance form, with generate_audio switched on.

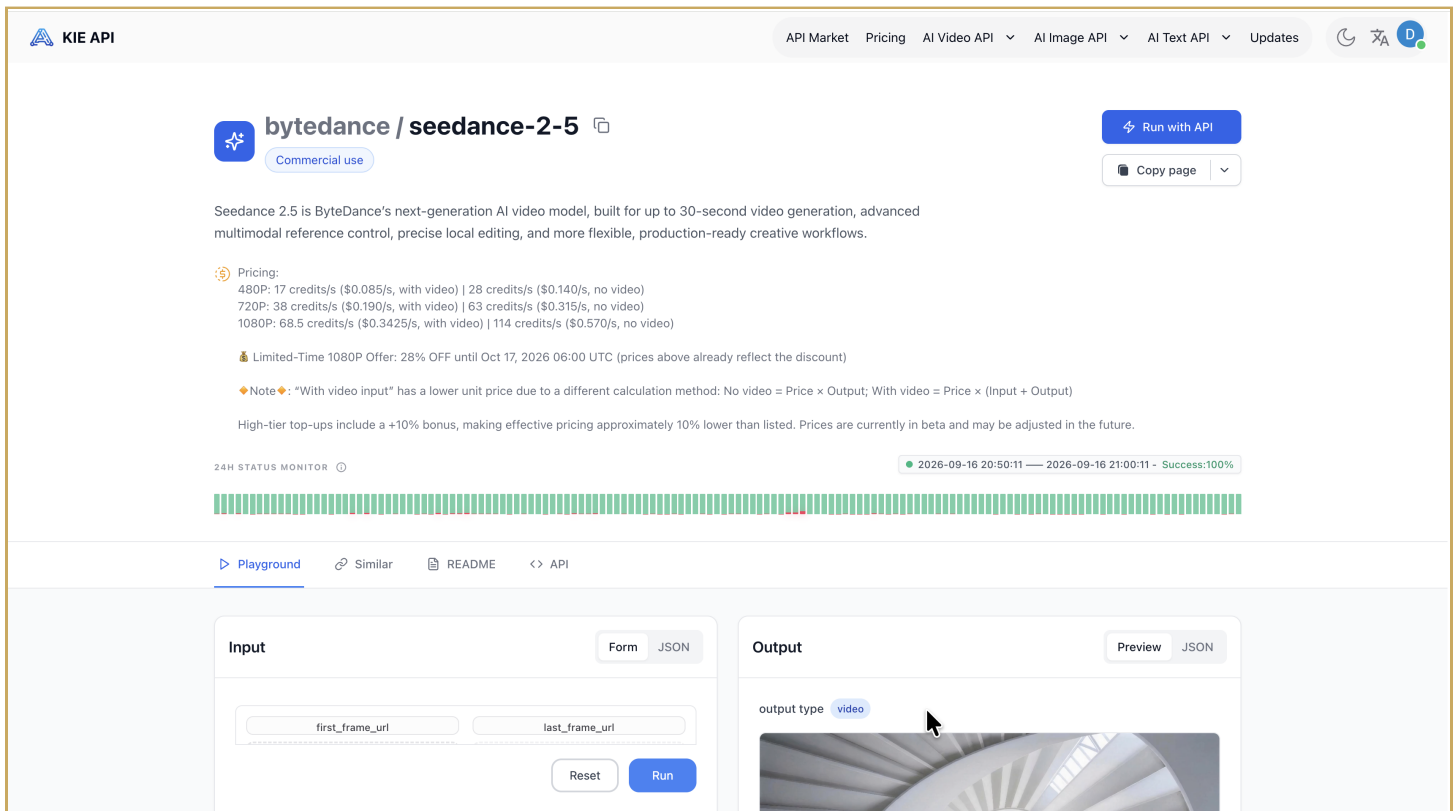
SCREENSHOT 19



A voice track loaded and ready — Kie.ai will generate the video synced to this audio.

The same run, screenshot by screenshot

SCREENSHOT 8



Seedance 2.5 selected on Kie.ai, pricing and Playground visible.

SCREENSHOT 9

KIE API

API Market Pricing AI Video API AI Image API AI Text API Updates

Playground Similar README <> API

reference_image_urls Remove All

File 1 Remove

File 2 Remove

+ Add more files (2/30)

A list of input image URLs.

reference_video_urls Remove All

Reset Run

Clearing the demo placeholder image, video, and audio files the form ships with by default — easy to miss, and they'll ruin the generation if left in.

SCREENSHOT 10

KIE API

API Market Pricing AI Video API AI Image API AI Text API Updates

Playground Similar README <> API

+ Add more files (1/10)

A list of input audio URLs. Furthermore, the total length of the three audios must not exceed 30 seconds.

generate_audio Whether to generate AI audio synchronized with the video. ☒

return_last_frame Whether to return the last frame of the video. When draft=true, this parameter cannot be set to true. ☐

resolution 480p 720p 1080p

The output video resolution.

aspect_ratio 16:9 4:3 1:1 3:4 9:16 21:9 adaptive

The aspect ratio of the generated video.

duration 30 x

Video duration in seconds.

output_format mp4 mov

Video output format.

web_search

Reset Run

Resolution set to 720p, aspect ratio to 9:16, duration to 30 seconds.

SCREENSHOT 11

KIE API

API Market Pricing AI Video API AI Image API AI Text API Updates

Playground Similar README <> API

their own version of this never works — pass.

The text prompt or description for the video. 8328/30000

reference_image_urls Remove All

File 1 Remove

+ Add more files (1/30)

A list of input image URLs.

reference_video_urls

Click to upload or drag and drop

Supported formats: MP4, QUICKTIME, X-MATROSKA Maximum file size: 200MB; Maximum files: 10

Select files

Reset Run

The real product reference image uploaded — the step it's easiest to forget, done before Run gets pressed.

SCREENSHOT 12

KIE API

API Market Pricing AI Video API AI Image API AI Text API Updates

Playground Similar README <> API

Video output format.

web_search Enable online search? ☐

nsfw_checker A configurable parameter. Defaults to true in the Playground. ☒

There is no guarantee that everything can be filtered out; if you are not satisfied with the results, you will need to make your own arrangements.

Reset Run

Everything set — hit Run.



A frame from the finished clip — product held steady, mic in shot, nothing warped.

Step 4a — get Claude to check it before you trust it

Don't eyeball a 30-second clip once at normal speed and call it done — AI video fails in small, specific ways that are easy to miss unless you're looking for them. The fastest way to catch them is to hand the clip straight back to Claude and let it do the checking.

1. **Download the finished clip from Kie.ai** and open a chat with Claude — the same one that wrote the prompt, or a fresh one, either works.
2. **Upload the video file directly and ask Claude to review it against the checklist** from Step 4: product continuity (same size, shape, and orientation throughout), hand rendering (five fingers, no warping), whether anyone walks through the background, face consistency start to finish, microphone continuity, and whether the subject walks past camera on the close instead of turning back.
3. **Claude works through the clip and tells you, in plain terms, what passes and what doesn't** — for example "the product looks like it warps slightly around 14–15s" or "there's a pedestrian visible in the background at 22s." This is the same checklist run against the finished Halden clip in Module 6.
4. **If something's off, don't start over from scratch.** Ask Claude which specific line in the prompt is most likely causing it — a missing continuity rule, a beat that's rushed, a direction that's too literal — and let it rewrite just that section, the same way you revised the dialogue and location earlier in this chat.

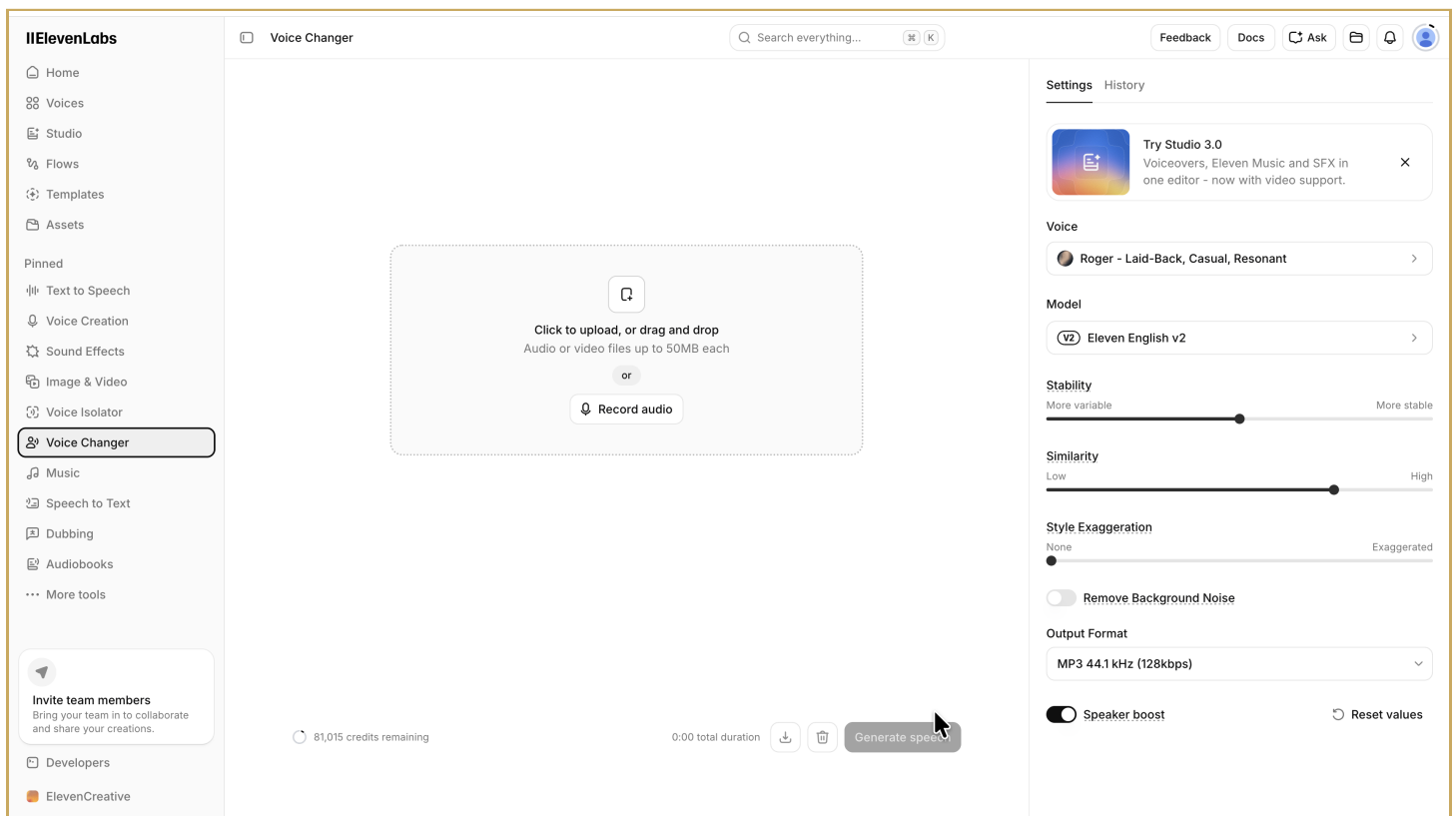
5. **Take the revised prompt back to Kie.ai and hit Run again.** Because the reference image and every other setting stay the same, this is usually a smaller, targeted fix rather than a full re-roll — but it's still a fresh generation and costs the same credits as any other run, so it's worth being specific about what needs to change before you regenerate rather than guessing.
6. **Repeat until it passes.** Most usable clips take two or three passes. If the same problem keeps showing up across multiple generations — a background pedestrian, say — that's almost always fixed by making the relevant line in the prompt more explicit, not by regenerating blind and hoping the next roll is cleaner.

Step 4b — Clone or swap the voice with ElevenLabs

Even when everything else about a generation is dialed in, the voice is often the last remaining tell. ElevenLabs' Voice Changer (Speech to Speech) takes the audio straight out of your finished clip and swaps in a different voice while keeping the exact original timing, pacing, and pauses — so the new audio drops back onto the video with no manual re-sync needed.

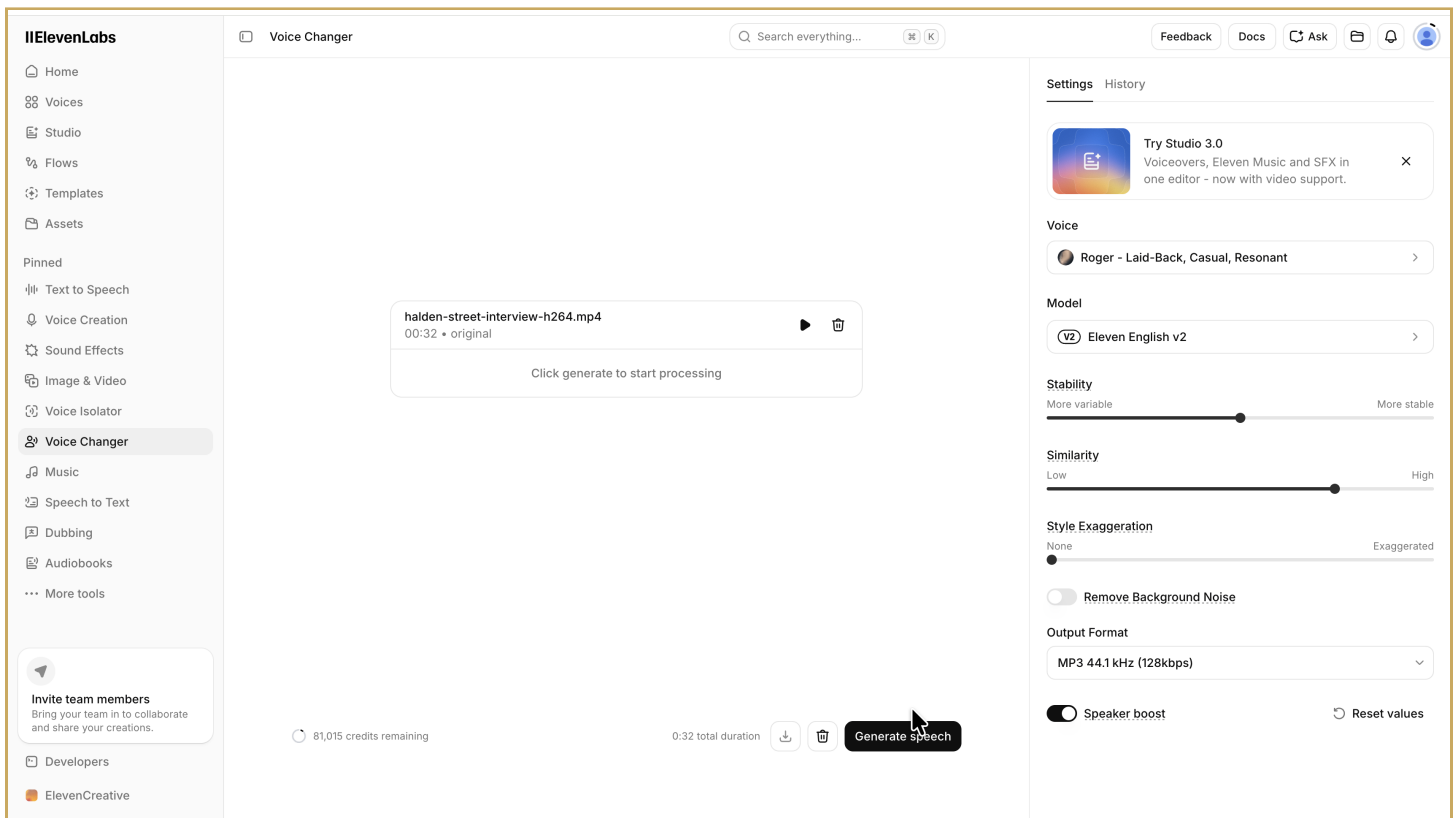
1. Go to **elevenlabs.io** and open **Voice Changer** from the tool tabs along the top (next to Speech, Image, Video, and Sound Effects).
2. **Drop your finished clip straight in.** It accepts audio or video files up to 50MB, so the exported .mp4 from Step 4 works as-is — there's no need to pull the audio out separately first.
3. **Pick a voice.** Search the Explore library by tone (try "casual conversational," "warm," or similar) for a natural-sounding option, or open My Voices to clone your own from a short reference recording if you want a specific person's voice.
4. **Watch for the fine-tuning warning.** Some library voices flag "not fine-tuned for Speech to Speech" when selected — those can sound unstable on this specific tool. Voices built for Speech to Speech (like the tool's own default) are the safer, more reliable pick.
5. Hit **Generate speech.** ElevenLabs re-renders the same words in the new voice while preserving the original delivery, so mouth movement in the source clip still lines up closely enough for social.
6. **Download the result** and swap it in during Step 5 below — mute the original audio track in CapCut and lay this one underneath in its place.

SCREENSHOT 14



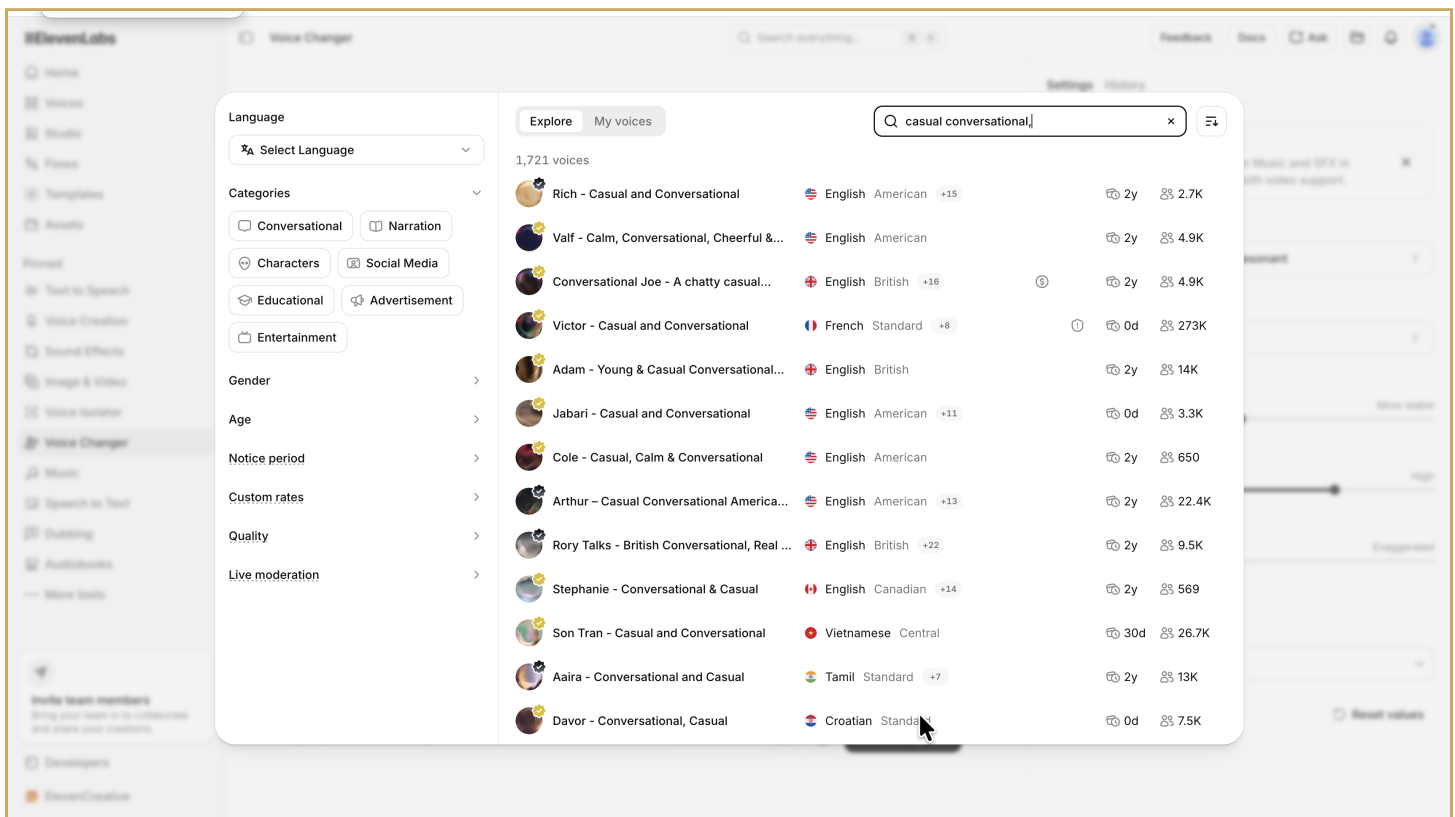
Voice Changer open, ready for a file — the default voice shown here (Roger) is one of the safer picks since it's built for Speech to Speech.

SCREENSHOT 15

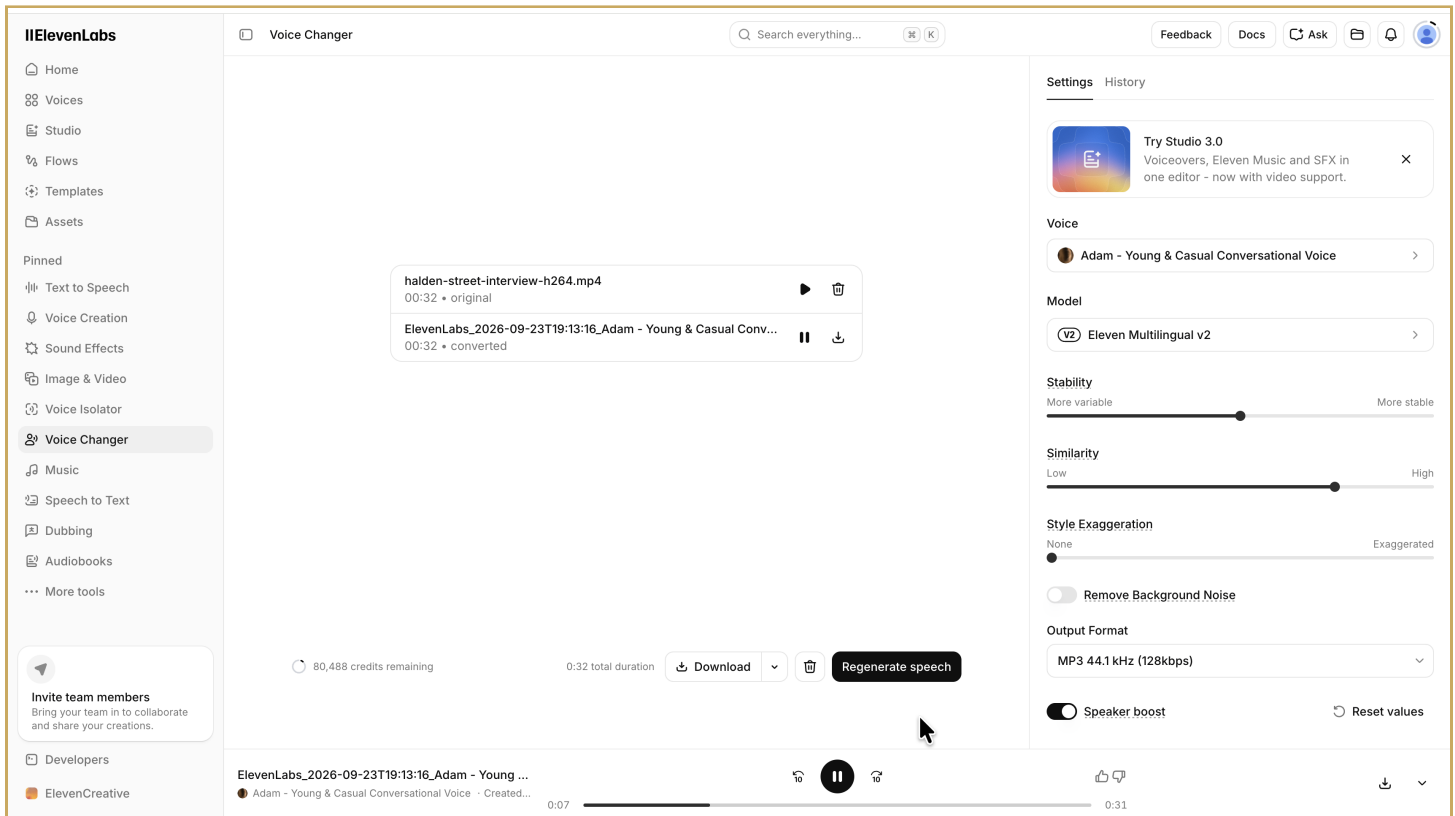


The finished clip dropped straight in — video and all, no need to pull the audio out first.

SCREENSHOT 16



Searching the Explore library by tone — "casual conversational" turns up hundreds of options to audition.



The converted result, ready to download and drop into CapCut in place of the original audio track.

Step 5 — Light post-production

No captions, no color grading, no music are generated by the AI on purpose — that's deliberately left to you, because AI-generated captions and grading tend to look worse than doing it yourself in CapCut in two minutes. Add burned-in captions, a light color pass if you want the "graded down to sell the flip" look, and you're ready to post.

Optional finishing touch — an end card. A two-second still at the very end, product name plus where to buy over a plain background, gives viewers a clean beat to land on and a second chance to register the name. It's a two-minute add in CapCut.



The actual end card from the Halden clip. That's the whole idea — it's not trying to do anything clever.

Step 6 — Post it or run it as a paid ad

This format works both organically (as native-feeling content on a brand or personal account) and as paid ad creative on TikTok and Meta, where its "not an ad" look is exactly what earns it a lower cost per click than polished creative.

5. The Prompt Library (the realism template)

This is the exact structure to copy for every street interview ad you generate. Fill in every bracket — the realism rules are what stop the AI model from drifting into "obviously AI" territory.

Format: 9:16 vertical, 30 seconds, one continuous take, no cuts — iPhone rear camera, handheld, operator standing ~1.5m from the subject.

The five structural rules that matter most

- 1. No one else walks through the background.** Moving background people are the single biggest cause of warped faces and duplicated bodies in AI-generated video.
- 2. Don't describe the product's label or packaging in words.** Let the reference image carry that — refer to the product only as "the bottle," "the packet," "the tin."
- 3. One physical action, one micro-expression, per beat** — not per sentence.
- 4. Restrained direction, warm words.** The subject can say "Oh wow, that's amazing" — the emotion lives in the words. The stage direction should say "eyebrows lift slightly," not "a big excited grin."
- 5. Overcast lighting by default.** Flat, diffused daylight kills the specular highlights on skin that make AI faces read as plastic.

The full fill-in-the-blank template — covering camera direction, lighting, setting, character description, the interviewer's mic and hand, the anti-glitch rules, product-continuity rules, and the "absolute no" list — is reproduced in full in the Appendix at the end of this kit.

6. A Real Worked Example, Start to Finish

Everything in this module is a genuine production, not a hypothetical — the actual prompt that generated it, the actual output, and the actual checks that were run on the finished file before it went anywhere near an ad account.

The product: an eau de toilette. **The benefit claimed:** bergamot and black pepper over cedar, still going strong at the end of the day. **The flip axis:** most aftershave fades by early afternoon — this one is still there.

A trademark lesson worth knowing before you start. The first version of this ad was generated against a real, well-known branded fragrance, using its actual bottle as the reference image. It was never used. The brand's wordmark was legible on the bottle and a liveried van appeared in the background — and an AI-generated stranger appearing to endorse a real commercial product he was never actually given is a fabricated endorsement, on top of the trademark problem of an unlicensed brand appearing in paid media. **If you don't own or have a licence for the brand you're advertising, invent one.** A plain, unbranded bottle with a name you made up costs you nothing and removes the risk entirely — the version below uses an invented brand, "Halden," for exactly this reason. This applies to more than fragrance: any product shot, sign, vehicle, or background detail that could read as a real company's branding needs to either be licensed or deliberately blanked out in the prompt.

The dialogue (91 words)

Time	Speaker	Line
0.0–2.0s	Interviewer	"Excuse me — can I ask you a question?"
2.0–3.0s	Subject	"Yeah, sure."
3.0–5.5s	Interviewer	"Be honest. Can you still smell your aftershave right now?"
5.5–13.5s	Subject	"Nah. Honestly, no. And I put it on this morning — proper few sprays, neck, wrists, the whole lot. By about eleven it's just gone. I've started carrying the bottle in my bag, which is a bit of a pain."
13.5–19.0s	Interviewer	"Have you tried Halden? Bergamot and black pepper over cedar, still going strong at the end of the day. Give it a go."
19.0–25.5s	Subject	(sprays wrist, brief pause) "Oh, that's nice. If that's still on me at six, I'm not buying anything else."
25.5–28.0s	Interviewer	"It's on Amazon — just search Halden."
28.0–30.0s	Subject	"Amazing. Ordering that one tonight."

Notice the shape: a genuine yes/no hook, one rambling true-to-life pain beat, the brand named clearly at the reveal and again at the close, and a decisive purchase-intent line to finish — with the bottle handed back to the interviewer as he turns to leave, which is what makes "ordering that one tonight" make sense instead of contradicting itself.



The hook — before he's noticed the interviewer, no microphone in shot yet.



The reveal at 16.2s — bottle held steady, mic in frame, caption timed to the line.

Watch the full 30-second clip this prompt produced:
d1yei2z3i6k35z.cloudfront.net/.../halden-street-interview-h264.mp4

What was checked on the finished file before it was considered usable

- **The product label** — cropped at full resolution and inspected to confirm the invented brand name was legible, undistorted, and free of any borrowed real-world design cues
- **The background** — every vehicle in frame checked for stray lettering or livery
- **Continuity of the take** — a silence-detection pass across the audio track, confirming continuous speech with no internal gaps, consistent with a genuine one-take recording
- **Frame-by-frame sampling** — ten evenly spaced frames pulled and checked by eye for face consistency, hand rendering, and product continuity

This is the actual quality bar to hold your own generations to — not "does it look okay at a glance" but "would it survive someone freeze-framing it." Two minutes of checking saves you from putting a warped hand or a flickering label into a paid ad campaign.

Post-production note: nothing about the video itself was changed — no recutting, retiming, or grading. The only processing was a file re-encode purely to get it under the file-size ceiling most ad platforms and gallery uploads enforce. Captions and the end card were added afterward in standard editing software, exactly as described in Module 4.

7. Posting & Ad Strategy

This format earns its place in your content mix two ways:

As organic content on a brand or personal account, it performs because it doesn't look like an ad in the feed — the "shaky, unpolished, someone got genuinely stopped" quality is what earns the watch time.

As paid creative on TikTok and Meta, it typically outperforms polished product shots on cost-per-click, because the ad-recognition part of a viewer's brain doesn't fire on it the way it does on a studio-lit product video. Run 3–5 variations with different flip axes against the same product and let the platform find your winner — this format is cheap enough to produce that testing multiple angles per product is the real unlock, not perfecting a single version.

8. The Community & Learning Hub

You don't have to figure this out alone. **Viral AI Ads Academy** is a paid community built around this exact style of AI UGC ad production — members trade prompts, troubleshoot generations that came out warped or robotic, post their finished clips for feedback before they go anywhere near an ad account, and share what's actually converting on Meta and TikTok right now.

It's the fastest way to get unstuck if a generation isn't landing, and a good place to see this format applied to products outside the ones covered in this kit. Optional, not required — everything you need to run this kit is already in the modules above.

\$27/MONTH COMMUNITY

Viral AI Ads Academy

Prompts, troubleshooting, and real finished clips from members running this exact system — shared as they post them, not filtered through a course.

See Viral AI Ads Academy →

Appendix — The Full Prompt Structure

This is the exact block the Master Prompt in Module 4 fills in for you — included here so you can see precisely what Claude is building, tweak the structure yourself, or build a prompt by hand without going through a chat tool at all. If you're following Module 4's workflow, you don't need to fill anything in here yourself — paste this whole block into the Master Prompt as instructed and Claude does the rest.

Battle-tested, generalised from a real production. Fill in every [BRACKET]. If your product doesn't involve a spray/cap action, delete the SPRAY MECHANICS and CAP CONTINUITY blocks — otherwise keep every block intact.

SEEDANCE 2.5 — STREET INTERVIEW, 9:16 VERTICAL, 30 SECONDS, ONE CONTINUOUS TAKE, NO CUTS

CAPTURE: iPhone 15 Pro rear camera, handheld by a person standing ~1.5m from the subject. Chest height, angled very slightly up, a degree or two off level. Constant handheld micro-shake from breathing and weight shifts. Natural drift and re-centring. No stabilisation, no zooms, no reframing, no cuts.

LOCATION: A wide UK high street, [TIME OF DAY]. Overcast, flat diffused daylight from directly above, no visible sun, no hard shadows. [3-5 unglamorous street details]. Every vehicle is a single flat colour with completely blank panels – no livery, no lettering, no logo, no stripe.

NOBODY WALKS PAST DURING THE TAKE. Background figures are stationary and distant only. No readable text and no brand identity anywhere in the background except the product itself.

FRAMING ARC (one take, operator movement only): [describe the walk-in, the stop, the medium-shot settle, the product hand-in at your reveal timestamp, and the walk-past exit – see Module 2's beat timings]

SUBJECT: [AGE, GENDER, ETHNICITY]. [Hair, skin texture, outfit – specific and unglamorous]. Eyeline on the interviewer standing just beside the lens, not into the lens itself.

INTERVIEWER: Off camera – voice, one hand, and a microphone. [Age, gender, tone]. A classic handheld reporter's stick microphone is raised into shot as the subject begins to slow, and stays in frame from then to the end. Only the hand and forearm are ever in shot. The product is offered with the OTHER hand.

ENERGY: [Narrow emotional arc. Never "delighted", "thrilled", "animated".]

SCRIPT: [Insert your five-beat script here, following Module 2's timings and Module 3's dialogue rules – one line of stage direction per beat: one physical action, one micro-expression.]

AUDIO: Raw iPhone mic outdoors. Two voices at natural conversational levels. Street ambience: traffic, distant chatter, occasional wind buffet on the mic. No music. No room tone. No stretch of total silence anywhere in the take longer than three quarters of a second.

DICTION (MANDATORY): every word spoken as written, in full, clearly intelligible; final consonants fully voiced; the brand name spoken exactly twice, identically both times, clearly audible.

ANTI-GLITCH & REALISM RULES (MANDATORY): natural blinking only (6-10 blinks); no face warping; five fingers on each hand at all times; hands anatomically correct during every gesture; lips sync naturally; hair and clothing respond naturally to movement; NO pedestrians walk past or cross frame at any point; vehicles move consistently and do not deform; subject's face and clothing remain identical start to finish; no beauty smoothing, no glow, no HDR sheen; no captions, no music, no jump cuts.

PRODUCT CONTINUITY (MANDATORY): [Product = the uploaded reference image. Reproduce its orientation exactly – never rotate, mirror, flip or invert the front face or its lettering.] Do not describe the packaging in the prompt. Refer to it only by a neutral noun ("the bottle", "the tin", "the packet"). The product is continuously visible during every handoff – no disappearing, no teleporting. Both hands fully rendered throughout, five fingers each. The product does not warp, bend, or change proportions when handled, and stays the same size relative to the hand throughout. Held steady and in sharp focus at the reveal, long enough to read. The handback at the close is a second handoff with the same rules.

SPRAY MECHANICS (MANDATORY, if applicable): the spray nozzle is a permanent part of the bottle and stays attached at every moment, including while the cap is off. The cap is a hollow cover with no mechanism – nothing ever comes out of the cap. The subject sprays by holding the BOTTLE upright and pressing the nozzle with a finger.

CAP CONTINUITY (MANDATORY, if applicable): the bottle is capped when it first enters frame. The cap is removed ON CAMERA, remains continuously visible once off, and is pushed back on ON CAMERA before the handback.

MICROPHONE CONTINUITY (MANDATORY): the microphone is NOT in frame when the take opens. It rises into frame between roughly 1.2s and 2.0s, before the subject speaks, and remains in frame continuously to the end. Classic handheld reporter's stick mic – handle plus rounded ball-shaped head, roughly 20-23cm, matte black or dark grey, unbranded. Never a long thin tube or shotgun mic.

ABSOLUTE NOS: no cinematic lighting; no smooth gimbal movement; no influencer pacing; no acting toward camera; no beauty filter; no colour grade; no static posture; no phone or camera visible as a prop; no boom pole or mic stand; no readable text anywhere in the background; no branding anywhere in the scene except the product itself; no vehicle livery; no shallow depth-of-field bokeh; no cuts, zooms, speed changes, or slow motion; no music or score; no interviewer's face, body, or shoulder in frame; no crowd reacting; no pedestrians crossing frame; no turning around and walking back the way they came; no real-world brand, logo, wordmark or product other than the one in the reference image.

FINAL TEST: if it feels like an ad, fail. If it feels like polished UGC, fail. If it feels like someone reading a script, fail. If he sounds like he already knows the brand, fail. If the product changes or warps in any frame, fail. If the brand name is mispronounced or differs between its two mentions, fail. If anyone walks past in the background, fail. If he walks off still holding the product, fail. If it feels like a stranger stopped on their lunch break who is quietly annoyed that their own version of this never works – pass.

That's the system.

Start with Module 4, generate your first clip today, and use Module 8 if you get stuck along the way. This format is only as good as the details — get them right and nobody watching will be thinking about how it was made.

kbvpublishing.com

viralaiads.kbvpublishing.com