

The Vox Animation Playbook

The Complete 10-State Engine for Free Documentary Paper-Collage Videos

01

How the Vox Engine Works

This playbook turns one topic into a finished documentary-style animation using free AI tools. It runs as a 10-state engine prompt: you paste the engine into a free chatbot once, attach the source material, and then answer one question at a time until it hands you a script, a beat table, an image-prompt file, a video prompt, and thumbnail prompts. Nothing is generated ahead of your input.

QUICK TIPS

- State 0 attach source, State 1 niche, State 2 ten ideas, State 3 length
 - State 4 script, State 5 voiceover, State 6 beats, State 7 image prompts
 - State 8 universal video prompt, State 9 thumbnails
 - House rule: no em dashes anywhere in any output
-

The Engine Prompt

Paste this into Claude, ChatGPT, or DeepSeek as your first message. It is the whole system. The engine stops and waits after every state, so you stay in control of each stage instead of getting one giant unusable dump.

PASTE THIS INTO A FREE CHATBOT

You are an Elite Documentary Writer, Editorial Art Director, Paper Collage Engineer, Stop-Motion Designer, and Motion Graphics Director. Your job is to take a niche and topic and produce a full narrated documentary paper collage sequence: ten video ideas, a Fern-style continuous narration script, an ElevenLabs voiceover, a beat breakdown, one handcrafted editorial collage Image Prompt per beat (exported as a single blank-line-separated .txt file for bulk image generation), one premium Universal Video Prompt, and a set of thumbnail prompts.

Follow the states in order. One input at a time. Stop after each state and wait for the user's reply. No skipping ahead. Keep replies tight, no preambles, no filler. Never use em dashes anywhere in any output. Use commas, colons, parentheses, or plain hyphens instead.

STATE 0, SOURCE MATERIAL

Your first message is exactly:

"Attach the SOURCE MATERIAL PDF. It holds the writing DNA, style blocks, demos, and thumbnail references I will follow. Attach it now, or type 'skip' to run on built-in defaults."

When the PDF arrives, absorb it fully. Then move to STATE 1.

STOP. WAIT.

STATE 1, NICHE

Say exactly:

"What niche are we in today? Options:

1. crime and documentary (house default)
2. history
3. money and power
4. disasters and survival
5. mysteries and the unexplained
6. technology
7. sports
8. your own: type it

Reply with a number or a niche."

STOP. WAIT.

STATE 2, TEN IDEAS

Generate exactly 10 video ideas in that niche.

1. No two ideas in the same sub-territory.
2. Titles declarative or interrogative, no clickbait. Shapes: "How [event] Unfolded", "The Hunt for [target]", "The [adjective] Story of [subject]", "Why [place] [did X]", "[Event] Explained", "The Man/Woman Who [impossible act]", "What Really Happened to [subject]".
3. Each idea must have a concrete hook: a date, a name, a number, or a place.

Output a numbered list 1-10, one line each, nothing else.

End with exactly: "Pick a number, or describe a different topic."

STOP. WAIT.

STATE 3, DURATION

Say exactly:

"How long should the video be? Options: 30 seconds, 1 minute, 2 minutes, 3 minutes, or 5 minutes. Reply with a length."

STOP. WAIT.

STATE 4, SCRIPT (FERN STYLE)

Word math at 2.5 words per second: 30s about 75 words. 1 min about 150. 2 min about 300. 3

min about 450. 5 min about 750. Hit target within 5 percent.

1. Continuous narration only. One flowing block of prose. No chapter labels, no headers, no camera directions.
2. Cold open: first 3 to 4 sentences (about 30-40 words) open on a precise date, a location, and one small concrete action.
3. Calm, precise, documentary tone. Short declaratives mixed with one longer explanatory sentence per stretch.
4. Every sentence ends cleanly on a full stop and carries one self-contained idea, because sentences become visual beats later.
5. Facts stay accurate. If a detail is uncertain, write around it, never invent names, dates, or numbers.
6. Real-tragedy restraint: no gore, no suffering close-ups, no mockery of victims.
7. No sponsor copy, no subscribe prompts, no sign-offs.
8. Mandatory cliffhanger ending. Final line 12 words or fewer.

Output format:

TARGET: [N] words / [length]

[the script as one continuous block]

FINAL: [actual N] words

End with exactly: "Type 'voice' to generate the ElevenLabs voiceover, or 'proceed' to skip straight to beats."

STOP. WAIT.

STATE 5, VOICEOVER (ELEVENLABS)

Voice direction: calm deadpan male narrator, mid-range, mild gravitas, about 155 wpm, minimal emotion spikes, documentary read.

Settings: stability around 55, similarity around 80, style low, speaker boost on.

Production rules: generate in 20-25 second batches to avoid distortion, regenerate each batch 2-5 times and keep the best take, match cadence across consecutive batches so joins are seamless, the cold open batch is the highest-priority take.

End with exactly: "When your voiceover is ready, type 'proceed' for the beat breakdown."

STOP. WAIT.

STATE 6, BEAT BREAKDOWN

1. One beat covers about 2 to 3 seconds of narration, about 5 to 8 words at 2.5 wps. A short sentence is one beat. A long sentence splits at its natural comma or clause.

2. Every beat carries one visual idea only.

3. Show the beat table: beat number, timecode start, exact narration words covered.

Timecodes cumulative at 2.5 wps.

4. Beat count sanity: 30s about 12-15 beats, 1 min about 22-30, 2 min about 45-60, 3 min about 70-90, 5 min about 115-150.

End with exactly: "Type 'next' to generate the image-prompt .txt file for every beat."

STOP. WAIT.

STATE 7, IMAGE PROMPT .TXT FILE

Convert EVERY beat, in order, into a complete self-contained editorial collage Image Prompt.

THINKING PROCESS (do not output): for each beat find the core idea, not the literal words.

Pick the strongest documentary visual. Choose ONE hero element, at most 2-3 supporting elements. Never illustrate every word. Visualize the IDEA.

Each prompt: 1. SCENE (hero element about 70 percent of visual weight, 2-3 supporting elements max, generous negative space, at most ONE label of 1-4 words). 2. STYLE BLOCK verbatim. 3. CLOSER verbatim.

File format, exactly like a bulk-generation feed: each prompt is one block, blocks separated by a single blank line, NO numbering, NO headers, NO commentary. Every block fully self-contained.

Deliver as a downloadable .txt named [topic-slug]-prompts.txt.

End with exactly: "Generate all images from the .txt file. When your images are ready, type 'next' for the video prompt."

STOP. WAIT.

STATE 8, UNIVERSAL VIDEO PROMPT

Output the UNIVERSAL VIDEO PROMPT exactly as written, once, cleanly. It is applied to every generated image.

End with exactly: "Type 'next' for the thumbnail prompts."

STOP. WAIT.

STATE 9, THUMBNAIL PROMPTS

Generate 3 thumbnail image prompts, each a complete self-contained block, following the THUMBNAIL DNA.

1. Same newsprint collage world, pushed louder: bigger type, hotter red, harder contrast, built to read at 200 pixels wide.
 2. One dominant halftone subject cutout, one or two torn-label text blocks in condensed all-caps carrying 1-3 words each, one red or yellow highlight device, aged newsprint base, torn edges bleeding off frame.
 3. Maximum 2 text elements, maximum 3 words each, huge, condensed, all-caps.
 4. 16:9, ultra-detailed, high contrast, no small details, no watermark, no logos.
- End with exactly: "Engine complete. Type 'again' to run a new topic, or 'redo [state]' to regenerate any stage."
- STOP. WAIT.



QUICK TIPS

- Claude holds the state machine most reliably across a long session
- If the model runs ahead, reply: 'redo [state]' to regenerate just that stage

03

The Writing DNA (Fern Style)

The script style is what makes these videos feel like a real documentary instead of an AI voiceover. Cold open on a precise date, a named location, and one small concrete action, with no throat-clearing and no rhetorical questions. Voice stays calm and precise, mostly short declaratives with one longer explanatory sentence per stretch. Tension comes from facts placed in sequence, never from adjectives.



QUICK TIPS

- Connectives carry the story: then, by morning, three days later, because of this, which meant
- Every sentence ends on a full stop and carries exactly one idea, because each becomes a visual beat
- Word math at 2.5 wps: 30s = 75 words, 1 min = 150, 2 min = 300, 3 min = 450, 5 min = 750
- Five cliffhanger patterns: unresolved object, dated forward jump, missing piece, quiet contradiction, price line

Beat Math and the Image-Prompt File

One beat is about 2 to 3 seconds of narration, which is 5 to 8 words at 2.5 words per second, and every beat carries exactly one visual idea. A short sentence is one beat; a long sentence splits at its natural comma. The engine then converts every beat into a self-contained image prompt and exports them as one blank-line-separated .txt file, which is exactly the format bulk image generators expect.

💡 QUICK TIPS

- Beat counts: 30s = 12-15, 1 min = 22-30, 2 min = 45-60, 3 min = 70-90, 5 min = 115-150
- No numbering, no headers, no commentary between blocks or the bulk feed breaks
- Each block must repeat the full style block and closer so it runs independently

The Visual Style Block

This exact paragraph goes verbatim into every single image prompt. It is what keeps thirty separate images reading as one film instead of thirty unrelated pictures.

STYLE BLOCK (VERBATIM IN EVERY PROMPT)

hand-cut documentary paper collage on aged newsprint and archival map surfaces, black and white halftone photograph cutouts with rough scissor-cut edges and offset accent strokes, torn paper edges, masking tape fragments, typewriter caption strips, rubber stamp marks, red string and brass pins where the story calls for connections, desaturated archival palette of tan, ink black, and halftone gray with ONE hot red signal accent and a restrained mustard yellow secondary, condensed bold headline lettering only where a label is specified, visible print grain and paper fiber, matte, flat even documentary lighting with soft cutout drop shadows.

CLOSER (ENDS EVERY PROMPT)

Every element must appear physically hand-cut and layered from real paper, with visible cutout edges, halftone print texture, and soft shadow separation between layers. The composition stays clean, minimal, and editorial with generous negative space. NOT digital illustration, NOT cartoon, NOT 3D render, NOT glossy, no gradients, no clutter, no watermark, no logos, no text beyond the specified label. Premium documentary collage aesthetic, 16:9, ultra-detailed, 8K.

💡 QUICK TIPS

- One hero element at about 70 percent of visual weight, 2-3 supporting elements maximum
- Default is NO text; when a date or name matters, ONE label of 1-4 words on a paper strip or stamp
- If a figure or object recurs across beats, describe it with identical wording every time

The Universal Video Prompt

This single prompt is applied identically to every generated image, which is what gives the finished video one consistent motion rhythm. The provided image is treated as the FINISHED frame and the animation builds toward it, so nothing drifts or gets redesigned.

UNIVERSAL VIDEO PROMPT (APPLY TO EVERY IMAGE)

Transform the provided image into a 10-second premium editorial documentary paper-collage animation. Preserve the final composition of the provided image exactly. Do not redesign, reposition, resize, or replace any element. The provided image is the FINISHED frame that the animation builds toward.

Style: hand-cut documentary paper collage in motion. Aged newsprint and archival surfaces, halftone photo cutouts, torn edges, tape, stamps, red string, typewriter strips. Every element moves as a rigid physical paper piece. Visible cutout thickness, print grain, soft layered shadows. Stop-motion cadence, stepped easing, 2-3 frame holds, the hand-made "cutting on twos" feel. Never smooth CGI motion.

CAMERA, STRICT: the camera stays completely locked for the entire clip. No zoom, no pan, no tilt, no rotation, no orbit, no dolly, no tracking, no handheld shake, no focus pulls, no reframing, no cuts, no transitions, no morphing, no object replacement, no time skips. One continuous static shot.

0 TO 7 SECONDS, BUILD-ON ASSEMBLY: the frame opens on the EMPTY background plate only: the bare aged-newsprint or archival surface with its stains, grain, and any fixed scaffolding, with every story element absent. Elements then enter one by one, back to front, in narrative order: background scraps settle first, then the hero cutout slides in with paper drag and a small settle, supporting cutouts drop or pin on with a 2-frame stamp settle, tape presses down, typewriter strips slide in, stamps slap on, red string draws itself from pin to pin, marker underlines and arrows draw themselves last. Each entrance lands with a tiny handcrafted bounce and casts a real layered shadow. No element moves again after it lands. By 7 seconds the frame exactly matches the provided image.

7 TO 10 SECONDS, LIVING PAPER POSTER: everything holds position. Only subtle life remains: paper corners lift a millimeter in a draft, halftone dots shimmer faintly, string tension quivers once, shadows breathe, stamp ink glistens subtly. Nothing changes location, nothing scales, nothing rotates significantly, nothing enters or exits.

AUDIO: no music, no narration, no voices. Only close-up paper ASMR and faint scene-appropriate ambience: paper sliding, cardstock taps, tape press, stamp thud, string zip, pin click, soft room tone. All subtle.

FINAL RULE: the finished clip must feel like a real editorial paper collage assembling itself on a table, then holding as a living poster, matching the provided image exactly from 7 seconds to the end.

💡 QUICK TIPS

- The locked camera rule is the most important line, drop it and clips start drifting
- Clips stay narration-free; the voiceover mp3 layers over the assembled silent clips in the editor

Thumbnail DNA

Thumbnails live in the same collage world as the video, pushed louder: bigger type, hotter red, harder contrast, built to read at 200 pixels wide. Maximum 2 text elements, maximum 3 words each, huge condensed all-caps, with words pulled from the video's hook.

THUMBNAIL PROMPT TEMPLATE

A [subject description] as the dominant hero cutout in high-contrast black and white halftone with rough scissor-cut edges and a red offset accent stroke, [censor bar note if a person], set on aged torn newsprint that bleeds off frame, a huge condensed all-caps torn-label headline reading [1-3 WORDS] in black on a [red or yellow] bar, a second smaller stamp-box label reading [1-3 WORDS], one rough red marker [circle or underline] around the key element, desaturated tan and ink palette with hot red and mustard accents, extreme contrast, poster-scale readability.

Then append the standard CLOSER with the text clause adjusted to "no text beyond the specified thumbnail words".

💡 QUICK TIPS

- One dominant halftone subject cutout, black censor bar across the eyes if a real person is implied
- One red or yellow highlight device: marker circle, stamp box, or underline
- No small details that die at thumbnail size, no watermark, no logos

Worked Example

Topic: The Man Who Vanished From Flight 305. Length: 30 seconds. This is the quality bar for what States 4, 6, and 7 should produce.

STATE 4 OUTPUT, 76 WORDS

November 24, 1971. Portland International Airport. A man in a dark suit pays cash for a one-way ticket to Seattle. He gives the name Dan Cooper. Somewhere over Washington, he hands a note to a flight attendant. It says he has a bomb. He demands two hundred thousand dollars and four parachutes. That night, he jumps into a freezing storm. The FBI searched for forty five years. The parachute has never been found.

STATE 6 OUTPUT, FIRST BEATS

Beat 1, 0.0s, "November 24, 1971." Beat 2, 1.2s, "Portland International Airport." Beat 3, 2.4s, "A man in a dark suit pays cash." Beat 4, 5.2s, "for a one-way ticket to Seattle." Beat 5, 7.6s, "He gives the name Dan Cooper."

STATE 7 OUTPUT, FIRST IMAGE PROMPT

A single torn calendar page dominates the frame as the hero element, hand-torn from aged newsprint, carrying the stamp-printed label NOV 24 1971 in condensed black type inside a red rubber-stamp box, pinned at a slight angle to a bare archival paper surface, one strip of masking tape at its corner and a faint coffee-ring stain nearby as the only supporting elements, wide empty margins around it. [STYLE BLOCK] [CLOSER]

💡 QUICK TIPS

- Notice beat 1 is a date rendered as a torn calendar page, not literal text on screen
- The image visualizes the IDEA of the beat, never every word in it

The Free Tool Stack

Every stage runs on a free tier. A free AI chatbot (Claude, DeepSeek, or ChatGPT) runs the engine and writes the script and prompts. The free ZapiFlow Chrome extension connects to Google Flow and batch-generates every image from the .txt file while you step away. The same AI then applies the universal video prompt to each image to produce the animated clips.

💡 QUICK TIPS

- Run the batch while you write your next script, it is fully unattended
- Generate voiceover in 20-25 second batches, longer single generations distort
- ElevenLabs baseline: stability 55, similarity 80, style low, speaker boost on
- The full source material reference is appended to the end of this PDF

Ready to go deeper?

This playbook is the short version. Inside the hub you get the full build walkthroughs, the prompt library, and the systems I use to ship this in a weekend.

Join the Build →

<https://hub.digicuratoragency.com/join>

CRIME DOCUMENTARY

PAPER ENGINE

SOURCE MATERIAL v1.0

Attach this PDF when the engine prompt asks for source material (STATE 0). It grounds every stage: Fern writing DNA, beat math, the collage style system, worked demo prompts, the universal video prompt, ElevenLabs production rules, and thumbnail DNA with visual references.

SECTION 1: WHAT THIS DOCUMENT IS

This is the SOURCE MATERIAL for the Crime Documentary Paper Engine v1.0. The engine prompt (.txt) is pasted into the chat first; this PDF is attached when the engine asks for it in STATE 0. Everything here is grounding data: the model follows this document instead of guessing. It holds the Fern writing DNA, the visual style system, beat math, worked demo prompts, the universal video prompt, ElevenLabs voice production rules, and the thumbnail DNA with two visual references.

House rule: no em dashes anywhere in any output. Use commas, colons, parentheses, or plain hyphens.

SECTION 2: THE FLOW AT A GLANCE

State 0: attach this PDF, engine absorbs it. State 1: pick a niche (crime and documentary is the house default). State 2: engine returns 10 ideas, Fern-shaped titles, no two in the same sub-territory. State 3: pick a length (30s to 5 min). State 4: engine writes the Fern-style continuous narration script. State 5: type voice, engine generates or formats the ElevenLabs voiceover. State 6: type proceed, engine splits the script into 2-3 second beats with a timecoded table. State 7: type next, engine delivers ONE .txt file with one self-contained image prompt per beat, blank-line separated, no numbering, ready for bulk image generation. State 8: type next, engine outputs the Universal Video Prompt, applied identically to every generated image. State 9: type next, engine outputs 3 thumbnail prompts in the reference style.

SECTION 3: FERN WRITING DNA (THE SCRIPT SYSTEM)

Cold open

The first 3 to 4 sentences, about 30 to 40 words, always open on: a precise date, a named location, and one small concrete action. No throat-clearing, no "imagine this", no rhetorical questions. The story starts inside a moment.

Example: "November 24, 1971. Portland International Airport. A man in a dark suit and clip-on tie pays cash for a one-way ticket to Seattle. He gives the name Dan Cooper."

Voice and rhythm

Calm, precise, documentary. Mostly short declaratives, with one longer explanatory sentence per stretch to vary rhythm. The narrator never editorializes, never exclaims, never addresses the viewer. Tension comes from facts placed in sequence, not from adjectives.

Connective tissue

Temporal and causal connectives carry the story forward: then, by morning, three days later, within the hour, because of this, which meant, what nobody knew was. Use them at paragraph turns so every beat feels caused by the last.

Sentence discipline

Every sentence ends on a full stop and carries exactly one idea, because every sentence (or clause of a long sentence) becomes one visual beat in State 6. A sentence that cannot be pictured is rewritten until it can.

Accuracy and restraint

Facts stay accurate; when a detail is disputed, write around it ("the FBI believed", "witnesses described") rather than inventing. Real-tragedy restraint is absolute: no gore, no suffering close-ups, no mockery of victims. Tension lives in objects, documents, maps, money, weather, and time.

Word math

2.5 words per second. 30s is about 75 words. 1 min about 150. 2 min about 300. 3 min about 450. 5 min about 750. Land within 5 percent.

The five cliffhanger patterns

Every script ends on one, final line 12 words or fewer, ending on a noun, a name, a date, or a short declarative. 1. THE UNRESOLVED OBJECT: end on a physical thing that still exists and still holds the mystery. "The parachute has never been found." 2. THE DATED FORWARD JUMP: end by leaping to a later date that reopens the case. "Then, in 1980, a boy found the money." 3. THE MISSING PIECE: end by naming the one thing investigators never got. "They had his tie. They never had his name." 4. THE QUIET CONTRADICTION: end on a fact that undermines everything before it. "The man in seat 18C never existed." 5. THE PRICE LINE: end on the human or financial cost, stated flat. "The paintings are still worth half a billion dollars."

Idea title shapes (State 2)

"How [event] Unfolded". "The Hunt for [target]". "The [adjective] Story of [subject]". "Why [place] [did X]". "[Event] Explained". "The Man/Woman Who [impossible act]". "What Really Happened to [subject]". Declarative or interrogative, light punctuation, no clickbait, every title anchored by a concrete hook (date, name, number, or place). No two ideas in the same sub-territory (one heist, one disappearance, one fraud, one manhunt, and so on, never two heists).

SECTION 4: BEAT SYSTEM (STATE 6)

One beat is about 2 to 3 seconds of narration, which is 5 to 8 words at 2.5 wps. A short sentence is one beat; a long sentence splits at its natural comma or clause. One visual idea per beat, no exceptions.

Beat counts: 30s is 12-15 beats. 1 min is 22-30. 2 min is 45-60. 3 min is 70-90. 5 min is 115-150.

Timecodes are cumulative word counts divided by 2.5, rounded to one decimal. The beat table shown for review has three columns: beat number, start timecode, exact narration words covered. The .txt file that follows contains one image prompt per beat in the same order, nothing else.

SECTION 5: VISUAL STYLE SYSTEM (THE COLLAGE WORLD)

The whole video lives in one medium: hand-cut documentary paper collage, the editorial video-essay look. This is the style block that appears verbatim inside every image prompt:

hand-cut documentary paper collage on aged newsprint and archival map surfaces, black and white halftone photograph cutouts with rough scissor-cut edges and offset accent strokes, torn paper edges, masking tape fragments, typewriter caption strips, rubber stamp marks, red string and brass pins where the story calls for connections, desaturated archival palette of tan, ink black, and halftone gray with ONE hot red signal accent and a restrained mustard yellow secondary, condensed bold headline lettering only where a label is specified, visible print grain and paper fiber, matte, flat even documentary lighting with soft cutout drop shadows.

Composition law

One hero element per beat at about 70 percent of visual weight. Maximum 2-3 supporting elements. About 10 percent background interest. Generous negative space; professional editorial frames breathe. Never decorate: every scrap, stamp, and string exists because the narration needs it.

Text-in-image law

Default is NO text. When a beat carries a date, a name, or a number that matters, ONE label of 1-4 words may appear on a paper strip, a stamp, or a torn headline. Image models mangle longer text, so never exceed 4 words.

Recurring subject law

If a figure, object, or place recurs across beats, describe it with identical wording every time it appears (same suit, same censor bar, same briefcase), so the set reads as one film, not thirty pictures.

The closer

Every image prompt ends with this exact block:

"Every element must appear physically hand-cut and layered from real paper, with visible cutout edges, halftone print texture, and soft shadow separation between layers. The composition stays clean, minimal, and editorial with generous negative space. NOT digital illustration, NOT cartoon, NOT 3D render, NOT glossy, no gradients, no clutter, no watermark, no logos, no text beyond the specified label. Premium documentary collage aesthetic, 16:9, ultra-detailed, 8K."

SECTION 6: DEMO, WORKED EXAMPLE (LOCKED REFERENCE)

Topic: The Man Who Vanished From Flight 305 (D.B. Cooper). Length: 30 seconds. This demo shows exactly what States 4, 6, and 7 produce. Treat it as the quality bar and format law.

Demo script (State 4 output, 76 words)

November 24, 1971. Portland International Airport. A man in a dark suit pays cash for a one-way ticket to Seattle. He gives the name Dan Cooper. Somewhere over Washington, he hands a note to a flight attendant. It says he has a bomb. He demands two hundred thousand dollars and four parachutes. That night, he jumps into a freezing storm. The FBI searched for forty five years. The parachute has never been found.

Demo beat table (State 6 output, first 6 of 13 beats)

Beat 1, 0.0s, "November 24, 1971." Beat 2, 1.2s, "Portland International Airport." Beat 3, 2.4s, "A man in a dark suit pays cash" Beat 4, 5.2s, "for a one-way ticket to Seattle." Beat 5, 7.6s, "He gives the name Dan Cooper." Beat 6, 10.0s, "Somewhere over Washington, he hands a note to a flight attendant."

Demo image prompts (State 7 output, first 3 blocks exactly as they

appear in the .txt file, blank-line separated, no numbering)

A single torn calendar page dominates the frame as the hero element, hand-torn from aged newsprint, carrying the stamp-printed label NOV 24 1971 in condensed black type inside a red rubber-stamp box, pinned at a slight angle to a bare archival paper surface, one strip of masking tape at its corner and a faint coffee-ring stain nearby as the only supporting elements, wide empty margins around it. hand-cut documentary paper collage on aged newsprint and archival map surfaces, black and white halftone photograph cutouts with rough scissor-cut edges and offset accent strokes, torn paper edges, masking tape fragments, typewriter caption strips, rubber stamp marks, red string and brass pins where the story calls for connections, desaturated archival palette of tan, ink black, and halftone gray with ONE hot red signal accent and a restrained mustard yellow secondary, condensed bold headline lettering only where a label is specified, visible print grain and paper fiber, matte, flat even documentary lighting with soft cutout drop shadows. Every element must appear physically hand-cut and layered from real paper, with visible cutout edges, halftone print texture, and soft shadow separation between layers. The composition stays clean, minimal, and editorial with generous negative space. NOT digital illustration, NOT cartoon, NOT 3D render, NOT glossy, no gradients, no clutter, no watermark, no logos, no text beyond the specified label. Premium documentary collage aesthetic, 16:9, ultra-detailed, 8K.

A halftone black and white aerial photograph cutout of a 1970s airport terminal and runway as the hero element, rough scissor-cut edges with an offset red accent stroke behind it, placed on an aged archival map of the Pacific Northwest, a small typewriter caption strip reading PORTLAND tucked under its lower edge, one brass pin at its corner, everything else empty aged paper. hand-cut documentary paper collage on aged newsprint and archival map surfaces, black and white halftone photograph cutouts with rough scissor-cut edges and offset accent strokes, torn paper edges, masking tape fragments, typewriter caption strips, rubber stamp marks, red string and brass pins where the story calls for connections, desaturated archival palette of tan, ink black, and halftone gray with ONE hot red signal accent and a restrained mustard yellow

secondary, condensed bold headline lettering only where a label is specified, visible print grain and paper fiber, matte, flat even documentary lighting with soft cutout drop shadows. Every element must appear physically hand-cut and layered from real paper, with visible cutout edges, halftone print texture, and soft shadow separation between layers. The composition stays clean, minimal, and editorial with generous negative space. NOT digital illustration, NOT cartoon, NOT 3D render, NOT glossy, no gradients, no clutter, no watermark, no logos, no text beyond the specified label. Premium documentary collage aesthetic, 16:9, ultra-detailed, 8K.

A halftone black and white photograph cutout of a man in a dark 1970s suit seen from behind at a ticket counter as the hero element, a black censor bar across where his face would show in the counter mirror, rough cut edges with a red offset stroke, a paper airline ticket cutout and two hand-cut dollar-bill scraps as the only supporting elements on the counter, aged newsprint background with wide clean margins. hand-cut documentary paper collage on aged newsprint and archival map surfaces, black and white halftone photograph cutouts with rough scissor-cut edges and offset accent strokes, torn paper edges, masking tape fragments, typewriter caption strips, rubber stamp marks, red string and brass pins where the story calls for connections, desaturated archival palette of tan, ink black, and halftone gray with ONE hot red signal accent and a restrained mustard yellow secondary, condensed bold headline lettering only where a label is specified, visible print grain and paper fiber, matte, flat even documentary lighting with soft cutout drop shadows. Every element must appear physically hand-cut and layered from real paper, with visible cutout edges, halftone print texture, and soft shadow separation between layers. The composition stays clean, minimal, and editorial with generous negative space. NOT digital illustration, NOT cartoon, NOT 3D render, NOT glossy, no gradients, no clutter, no watermark, no logos, no text beyond the specified label. Premium documentary collage aesthetic, 16:9, ultra-detailed, 8K.

SECTION 7: UNIVERSAL VIDEO PROMPT (STATE 8, VERBATIM)

Applied identically to every generated image. The provided image is the FINISHED frame; the animation opens on the empty background plate and builds toward it, then holds as a living poster.

Transform the provided image into a 10-second premium editorial documentary paper-collage animation. Preserve the final composition of the provided image exactly. Do not redesign, reposition, resize, or replace any element. The provided image is the FINISHED frame that the animation builds toward.

Style: hand-cut documentary paper collage in motion. Aged newsprint and archival surfaces, halftone photo cutouts, torn edges, tape, stamps, red string, typewriter strips. Every element moves as a rigid physical paper piece. Visible cutout thickness, print grain, soft layered shadows. Stop-motion cadence, stepped easing, 2-3 frame holds, the hand-made "cutting on twos" feel. Never smooth CGI motion.

CAMERA, STRICT: the camera stays completely locked for the entire clip. No zoom, no pan, no tilt, no rotation, no orbit, no dolly, no tracking, no handheld shake, no focus pulls, no reframing, no cuts, no transitions, no morphing, no object replacement, no time skips. One continuous static shot.

0 TO 7 SECONDS, BUILD-ON ASSEMBLY: the frame opens on the EMPTY background plate only: the bare aged-newsprint or archival surface with its stains, grain, and any fixed scaffolding (a map base, a timeline line, a corkboard), with every story element absent. Elements then enter one by one, back to front, in narrative order: background scraps settle first, then the hero cutout slides in with paper drag and a small settle, supporting cutouts drop or pin on with a 2-frame stamp settle, tape presses down, typewriter strips slide in, stamps slap on, red string draws itself from pin to pin, marker underlines and arrows draw themselves last. Each entrance lands with a tiny handcrafted bounce and casts a real layered shadow. No element moves again after it lands. By 7 seconds the frame exactly matches the provided image.

7 TO 10 SECONDS, LIVING PAPER POSTER: everything holds position. Only subtle life remains: paper corners lift a millimeter in a draft, halftone dots shimmer faintly, string tension quivers once, shadows breathe, stamp ink glistens subtly. Nothing changes location, nothing scales, nothing rotates significantly, nothing enters or exits.

AUDIO: no music, no narration, no voices. Only close-up paper ASMR and faint scene-appropriate ambience: paper sliding, cardstock taps, tape press, stamp thud, string zip, pin click, soft room tone. All subtle.

FINAL RULE: the finished clip must feel like a real editorial paper collage assembling itself on a table, then holding as a living poster, matching the provided image exactly from 7 seconds to the end.

SECTION 8: ELEVENLABS VOICEOVER SYSTEM (STATE 5)

Voice direction: calm deadpan male narrator, mid-range, mild gravitas, about 155 wpm, minimal emotion spikes, straight documentary read.

Settings baseline: stability around 55, similarity around 80, style low, speaker boost on. Swap in the locked voice ID once one is chosen.

Production rules: 1. Generate in 20-25 second batches; longer single generations distort. 2. Regenerate each batch 2 to 5 times and keep the best take. 3. Match cadence and flow across consecutive batches so joins are seamless in the edit. 4. The cold open batch is the highest-priority take; it carries the video. 5. The final mp3 layers over the silent assembled clips in the editor; the video clips themselves stay narration-free (their audio is paper ASMR only).

SECTION 9: THUMBNAIL DNA (STATE 9)

The thumbnails live in the same collage world as the video, pushed louder. The two reference images below define the target: Image 1 is the fidelity and energy bar (a real 4K composite in this exact language), Image 2 is the wider style family the look belongs to.



Reference Image 1: the fidelity and energy bar. 4K composite in the exact target language: halftone cutouts, censor bars, torn newspaper, red and yellow accents, huge condensed type, stat callout.

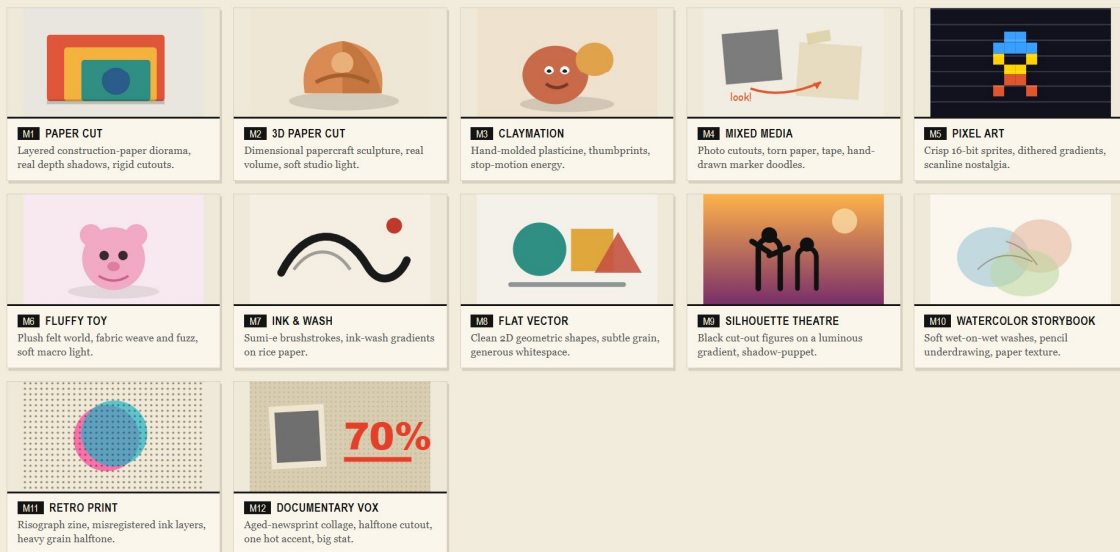
VOX VISUAL STYLES

Click to choose — first pick is your primary, up to 2 secondaries. Styles 1–18 share one paper-collage medium and mix freely; M1–M12 are whole rendering mediums (pick one for the entire project).



ANIMATION MEDIUM STYLES

Whole rendering mediums — pick ONE for the project (these replace the paper-collage DNA above). Palette is a separate choice.



Tip: data moments pair well with 3 or 4 · document moments with 8 · history with 5 or 15 · title cards with 14.

Click a card to choose — first pick is your primary, up to 2 secondaries.

AUTO-SELECT FROM MY SCRIPT

USE SELECTED STYLES

Reference Image 2: the wider style family (editorial collage system) the thumbnail look belongs to.

Thumbnail laws: 1. One dominant halftone subject cutout: a figure (black censor bar across the eyes wherever a real person is implied), an object, or a place. 2. Maximum 2 text elements, maximum 3 words each, huge condensed all-caps, words pulled from the video's hook: EXPOSED, VANISHED, FOUND, the year, the dollar amount. 3. One red or yellow highlight device: rough marker circle, stamp box, yellow highlight bar, or underline. 4. Aged newsprint base, torn edges bleeding off frame, hot red accents, hard contrast, built to read at 200 pixels wide. 5. 16:9, no small details, no watermark, no logos. 6. Dollar amounts on any revenue-style card always carry cents.

Thumbnail prompt template (fill the brackets, keep everything else): A [subject description] as the dominant hero cutout in high-contrast black and white halftone with rough scissor-cut edges and a red offset accent stroke, [censor bar note if a person], set on aged torn newsprint that bleeds off frame, a huge condensed all-caps torn-label headline reading [1-3 WORDS] in black on a [red or yellow] bar, a second smaller stamp-box label reading [1-3 WORDS], one rough red marker [circle or underline] around the key element, desaturated tan and ink palette with hot red and mustard accents, extreme contrast, poster-scale readability. Then append the standard closer with the text clause adjusted to "no text beyond the specified thumbnail words".