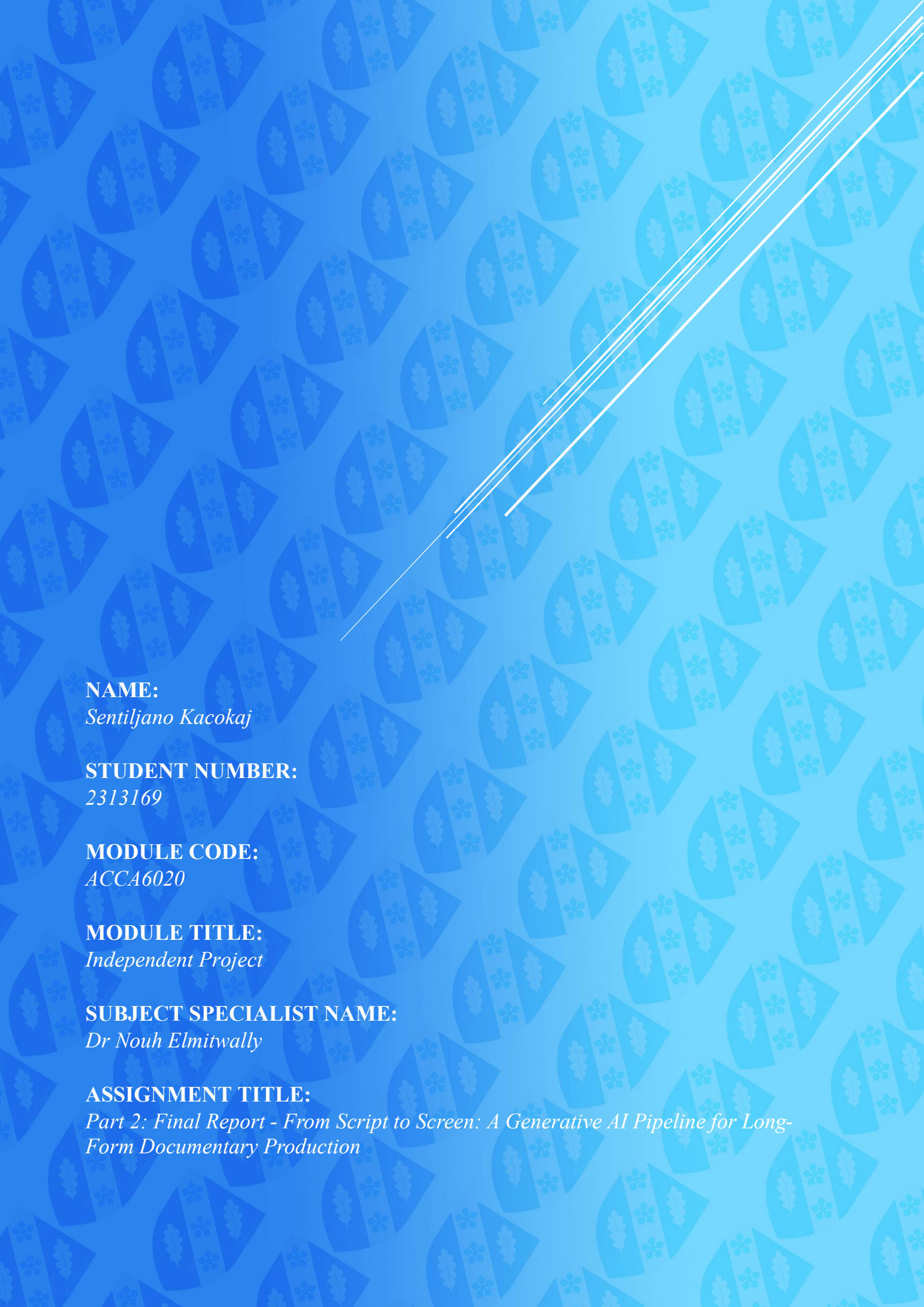




NAME:

Sentiljano Kacokaj

STUDENT NUMBER:

2313169

MODULE CODE:

ACCA6020

MODULE TITLE:

Independent Project

SUBJECT SPECIALIST NAME:

Dr Nouh Elmitwally

ASSIGNMENT TITLE:

Part 2: Final Report - From Script to Screen: A Generative AI Pipeline for Long-Form Documentary Production

I.	ABSTRACT	3
II.	INTRODUCTION	3
	A. Background	3
	B. Problem Statement.....	3
	C. Objectives.....	3
	D. Scope of the Project.....	4
	E. Dissertation Structure.....	4
III.	LITERATURE REVIEW.....	4
	A. Existing Research and Related Work.....	4
	B. Rationale and Research Gap.....	6
IV.	METHODOLOGY	6
	A. System Development Approach.....	6
	B. Tools and Technologies.....	7
	C. System Requirements	8
	D. System Architecture and Design	8
	E. Data Collection and Processing.....	10
	F. Ethical and Legal Considerations	10
V.	IMPLEMENTATION AND DEVELOPMENT	11
	A. The Prototype	11
	B. The Full Build on Self-Hosted Weights.....	11
	C. The V3 Pivot.....	12
	D. Video Selection and the Production Run.....	13
	E. Lesson Applied: Merged Beats Split	13
VI.	TESTING AND EVALUATION	13
	A. The Artifact and the Headline Numbers	13
	B. Efficiency.....	15
	C. Quality.....	15
	D. Robustness.....	16
	E. Limitations.....	16
VII.	CONCLUSIONS AND RECOMMENDATIONS	16
	A. Conclusions	16
	B. Recommendations and Future Work.....	17
VIII.	REFLECTION.....	18
IX.	REFERENCES	19
X.	APPENDICES	21
	A. The full shot sheet.....	21
	B. Credit accounting reconciliation	26
	C. Prompt system	26
	D. Identity reference sheets	29
	E. Final Gantt chart.....	30
	F. Artifact access	31

LIST OF FIGURES AND TABLES

LIST OF FIGURES

Figure 1. The three DSR cycles and what each carried forward.....	6
Figure 2. Pipeline architecture: three phases gated by the script lock, the identity lock and the frame lock.....	9
Figure 3. Unretouched V2 output.....	11
Figure 4. Frozen identity sheet next to the same character in scene.	12
Figure 5. Visual generation spend per finished minute against the indicative vendor baseline band, on a logarithmic scale.	15
Figure 6. The proposed orchestration screen, designed and deliberately unbuilt.	17
Figure A.1. Contact sheet of the final approved frame for each of the 58 shots, in shot order.....	21
Figure C.1. ComfyUI batch workflow for the stills pass without references, sending per-shot prompts to Nano Banana Pro through the hosted API.	27
Figure C.2. ComfyUI batch workflow for the stills pass with references. The column of thumbnails on the left holds the identity sheets and environment masters, routed to each generation node alongside its per-shot prompt.....	27
Figure C.3. The three-node video chain (LoadImage, ByteDanceImageToVideoNode, SaveVideo). The approved frame enters unchanged and the video prompt drives the middle node.....	28
Figure D.1. ComfyUI workflow for identity sheet generation. A single synthetic portrait drives two Nano Banana Pro passes, one full-body turnaround and one headshot turnaround, and a stitch node assembles the sheet that was then frozen.....	29
Figure D.2. Frozen identity reference sheet for Maxim Kuzminov in civilian clothing.	29
Figure D.3. Frozen identity reference sheet for Maxim Kuzminov in military flight attire.	30
Figure D.4. Frozen identity reference sheet for the two gunmen, body turnaround only, faces concealed in every view.	30
Figure E.1. The final Gantt chart. Ten task bars and seven milestones over 27 weeks, week 1 beginning 26 January 2026.	30

LIST OF TABLES

Table 1. The V3 stack at submission.	7
Table 2. Functional (F) and non-functional (N) requirements.	8
Table 3. The V2 to V3 swap, stage by stage.	12
Table 4. Headline results, derived from the shot sheet.	14
Table A.1. The full shot sheet, 58 rows, with attempts and credits per shot as logged during generation.....	25
Table B.1. Credit reconciliation between the account balance and the itemised production records.	26
Table C.1. The complete Prompt Writer V9 record for shot 0019, as delivered in shot_sheet_chunk_01 and filled during production.	26

I. ABSTRACT

Generative models now produce cinematic stills, natural narration and coherent scripts, yet a person still carries outputs between separate tools. This project asked whether one pipeline could take a documentary from a locked script to a finished segment, and what the recorded run would cost in money, time and regeneration events. Following Design Science Research, three pipeline versions were built and evaluated. The final version combines a four-stage script system, ElevenLabs narration, Nano Banana Pro stills held consistent by frozen character reference sheets, Seedance 1.5 Pro animation and CapCut assembly. Prior to being used downstream, approved assets are frozen by script, identity and frame locks. The artifact is a 3 minute 18 second documentary of 44 shots, cut from a 58-shot production run. Visual generation spend was \$9.56 per finished minute. Recorded animation and edit time was 73 minutes per finished minute, and regeneration events averaged 3.0. Published vendor estimates of \$1,000 to \$10,000 per finished minute provide an indicative conventional-production baseline because they include crew, equipment, travel and licensing. The success criterion, a clear advantage in one evaluation dimension, was met on efficiency.

II. INTRODUCTION

A. BACKGROUND

Affordable cameras changed who could film. They did not change how expensive documentaries were to produce. Published vendor estimates place conventional documentary production between \$1,000 and \$10,000 per finished minute [1], [2]. The figures provide an indicative industry baseline because they include crew, equipment, travel and licensing. Over the past four years the component tools of production evolved one by one. Diffusion models generate cinematic still images from text [3]. Neural speech synthesis produces narration that holds broadcast tone at length [4]. Large language models sustain structure over long-form writing [5]. Each performs well alone. In practice a person carries outputs between them by hand, reformatting, re-describing and re-uploading at every boundary, and the time the tools save is lost in the gaps between them.

Documentary reconstruction is a strong test case for closing those gaps. The form is narration-led, it tolerates stylised imagery where archival footage does not exist, and it needs no live subjects in front of a camera.

B. PROBLEM STATEMENT

Each tool in the chain has published benchmarks, but nothing equivalent exists for the joined workflow. No study reviewed for this project measures what happens when script, voice, image, video and assembly run as one pipeline with measurement built in: what a finished minute costs, where the process fails, and how often a person must step in. The problem this project addresses is that integration itself is unmeasured. The research question follows directly: can one pipeline take a documentary from a locked script to the screen, and at what cost in money, time and intervention?

C. OBJECTIVES

The proposal set a broader aim of an automated, episode-length workflow. Development narrowed the evaluated artifact to a 3 minute 18 second segment after the V3 visual pivot. The final build was assessed against four operational objectives:

1. Build a multi-stage script system that ends in a locked script.
2. Generate voiceover, images and video from the locked script.
3. Assemble the assets into a finished documentary segment.
4. Evaluate the production run using recorded visual generation spend, recorded animation and editing time, output quality and regeneration events.

The contribution of this project is the pipeline as a whole instead of any of the models inside it. The generative models are used as replaceable services. The design work sits in the three-lock pattern, the merged prompt system, the four batch workflows and the evaluation instrument that logged every shot as it was made.

D. SCOPE OF THE PROJECT

The evaluated artifact is "The Pilot Who Flew Home", a 3 minute 18 second reconstruction of the 2023 Kuzminov military defection. Its 44-shot cut was selected from a 58-shot production run. The artifact uses one synthetic lead and one archival photograph. Metrics are reported per finished minute and projected to episode scale in Testing and Evaluation. A full episode, live-action footage, real faces inside generative models and distribution remain outside the tested scope.

E. DISSERTATION STRUCTURE

The Literature Review surveys the component technologies and locates the research gap. The Methodology sets out the development approach, the architecture, the evaluation design and the ethical and legal position. Implementation and Development traces the build through three versions and two pivots. Testing and Evaluation reports the measured results and closes with limitations. The Conclusions and Recommendations section closes the argument and sets the forward work, and the Reflection looks back on the project as a piece of work.

III. LITERATURE REVIEW

A. EXISTING RESEARCH AND RELATED WORK

Script generation. A survey of large language models [5] covers instruction following and coherence over long outputs among their capabilities. Two of its observations shaped this project. Long-form quality degrades when a single prompt asks for everything at once, which argues for staged generation with review between stages. And model outputs vary between runs, which argues for freezing an approved text before anything downstream depends on it. The four-stage script system in the Methodology applies both. The staging also creates a natural place for editorial control. A documentary script is a chain of factual and tonal decisions, and a system that generates it in reviewed stages keeps a person at each decision without slowing the writing to human speed.

Speech synthesis. A systematic review of deep learning speech synthesis [4] reports that modern systems reach natural rhythm and intonation and keep one consistent voice at paragraph length, with expressiveness over very long inputs still an open problem. Documentary narration is a favourable case: one voice, a measured tone, no dialogue. Commercial systems expose this capability through hosted interfaces priced per character, which makes narration the cheapest and most stable stage of the pipeline. The remaining production risk is regeneration. A second run of the same text does not return the same performance, so an approved narration must be kept and never remade, and the pipeline treats the audio file the way it treats every other approved asset.

Image generation and the identity problem. Diffusion models synthesise images by iteratively denoising random noise under text conditioning [3]. They can produce photorealistic images on demand, but in production terms their weakness is that nothing carries over between runs. Each image is sampled fresh, so the same described character drifts in face, build and wardrobe between generations. For a documentary lead this drift is a serious problem, because the reconstruction stops holding together as a single story if the protagonist visibly changes between shots. Published work answers this in two ways. The first is fine-tuning: LoRA trains small low-rank adapter matrices on a handful of subject images while the base weights stay frozen, so the identity is baked into weights and recalled with a trigger token [6]. The cost is a training step, a dataset, GPU time and a fixed artifact that must be retrained to change. The second answer is reference-conditioned generation, in which the subject is supplied as an input image at generation time and the model conditions on it directly, with no training [7]. The distinction matters for production. Weight-based identity is stable but requires training infrastructure. Reference-based identity avoids that, though it depends entirely on preserving the reference file. Both approaches still depend on human judgement, just at different points in the process. Fine-tuning spends the judgement before training, in curating the subject dataset. Reference conditioning spends it at the moment the reference is approved, which is exactly where this project placed its identity lock. Implementation and Development shows the pipeline moving from the first family to the second, and the Methodology shows the frozen reference sheet doing the work the LoRA used to do.

Video generation. Image-to-video models animate a supplied still, conditioning the clip on that first frame so motion inherits the composition, palette and subject of the input [8]. First-frame conditioning is the mechanism that makes a frame lock possible at all: approve the still, and the animation is constrained to it. Motion is then steered with a short text instruction on top of the frame, and the instruction can stay plain because the composition is already decided; the prompt only has to say what moves. Hosted video models price per second of output, so cost is known before generation, a property the evaluation design in the Methodology depends on.

Generative AI in documentary practice. Synthetic media has already entered documentary, and the reception has been shaped by disclosure [9]. Practitioner debate has settled around three uses: fully synthetic reconstruction, restoration or enhancement of real archive, and synthetic voice or likeness of real people, with the third drawing the sharpest objections. The consistent finding is that audiences and practitioners accept reconstruction when it is labelled and reject it when it is passed off, and the ethics design in the Ethical and Legal Considerations section takes that as a requirement.

Evaluating generative systems. A review of generative AI evaluation [10] notes that output quality resists standard automatic metrics, while cost, latency and failure behaviour can be measured directly. That split shaped the evaluation design: quality is assessed by structured human review, and efficiency and robustness are logged as numbers during production. Its requirements framing also supports evaluating a system in use, under real workload, instead of on benchmark prompts.

B. RATIONALE AND RESEARCH GAP

Each strand above evaluates its own tool in isolation. Image papers report fidelity scores, speech papers report naturalness ratings, video papers report motion quality, and none of them measures the joined workflow those tools are supposed to serve. The practical questions a producer would ask are unanswered in the literature: the cost of a finished minute, the failure points between stages, the number of times a person must intervene, and whether a stage can be replaced without rebuilding the whole. The gap in the literature is therefore the integrated workflow itself, and measuring it requires production conditions instead of benchmark prompts.

This project fills that gap with a working pipeline and a per-shot evidence log. Documentary reconstruction was chosen because it exercises every stage, tolerates a synthetic visual style, and raises the ethical questions of synthetic media in a form that can be answered by design. Design Science Research fits a gap of this shape. The question concerns the behaviour of a built system under production load, so the method has to produce the system and then measure it; a survey or a component experiment would reproduce the very isolation the gap describes. The Methodology that follows sets this out, with the architecture and the evaluation design it carries.

IV. METHODOLOGY

A. SYSTEM DEVELOPMENT APPROACH

The project follows Design Science Research. The framework in [11] positions the built artifact as the research contribution, judged by evaluation against criteria fixed in advance, and [12] structures the work as a cycle: identify the problem, define objectives, design and develop, demonstrate, evaluate, communicate. The project ran that cycle three times. V1, in January, was a 1 minute 15 second prototype on the D.B. Cooper story that proved the three-phase structure. V2, February to June, was the full build on self-hosted model weights and was the position presented at the interim viva. V3, in July, rebuilt the visual branch on hosted reference-conditioned models. Each version was a complete build-and-evaluate pass, and the evidence for evaluation was logged during production, with every shot carrying its attempts, credits and interventions as it was made.

Figure 1 below shows the three DSR cycles and what each one carried forward.



Figure 1. The three DSR cycles and what each carried forward.

B. TOOLS AND TECHNOLOGIES

The V3 stack, current at submission, is set out in Table 1.

Phase	Tool	Role	Billing
1 Script	Claude	Four-stage script system ending in a locked script	Subscription
2 Voiceover	ElevenLabs V3, Oxley voice [13]	Narration generated once, at full script length	Per character
2 Stills	Nano Banana Pro [14]	Scene stills held consistent by frozen character reference sheets	Per image, in credits
2 Video	Seedance 1.5 Pro [15], called through ComfyUI as a thin API client [16]	Animation of every approved still	Per clip, in credits
3 Assembly	CapCut	Manual edit, archival insert and final credit labels	Free tier

Table 1. The V3 stack at submission.

Two properties drove selection. Hosted models bill per unit of output, so the cost of every asset is known exactly, a precision the self-hosted V2 stack never gave. And every stage exposes a plain interface, text or image in, asset out, so a stage can be swapped without touching its neighbours. Implementation and Development documents the swap that tested this in practice. The V2 stack that V3 retired is described there with its own tooling.

C. SYSTEM REQUIREMENTS

Table 2 states the requirements the pipeline was built against.

ID	Requirement
F1	Produce a locked script from a story brief through staged review.
F2	Generate narration once, at full script length, from the locked script.
F3	Generate a still for every scripted shot with one consistent lead identity.
F4	Animate every approved still.
F5	Assemble generated and archival assets with final credit labels.
F6	Log attempts, credits and interventions per shot during generation.
N1	Visual generation spend per finished minute at least an order of magnitude below the lowest indicative vendor baseline.
N2	No real person’s face enters any generative model.
N3	Any single stage replaceable without redesign of the others.
N4	The full evidence log reproducible from production records alone.

Table 2. Functional (F) and non-functional (N) requirements.

F6 and N4 make the evaluation part of the system instead of an afterthought; the Data Collection and Processing section explains how.

D. SYSTEM ARCHITECTURE AND DESIGN

The pipeline has three phases. Phase 1 produces a locked script through a four-stage script system with human review between stages. Phase 2 turns the locked script into assets: narration once at full length, then a still for every shot, then a clip from every approved still. Phase 3 assembles clips, statics and archival inserts against the narration in CapCut. The three phases and the position of each lock are shown in Figure 2 below.

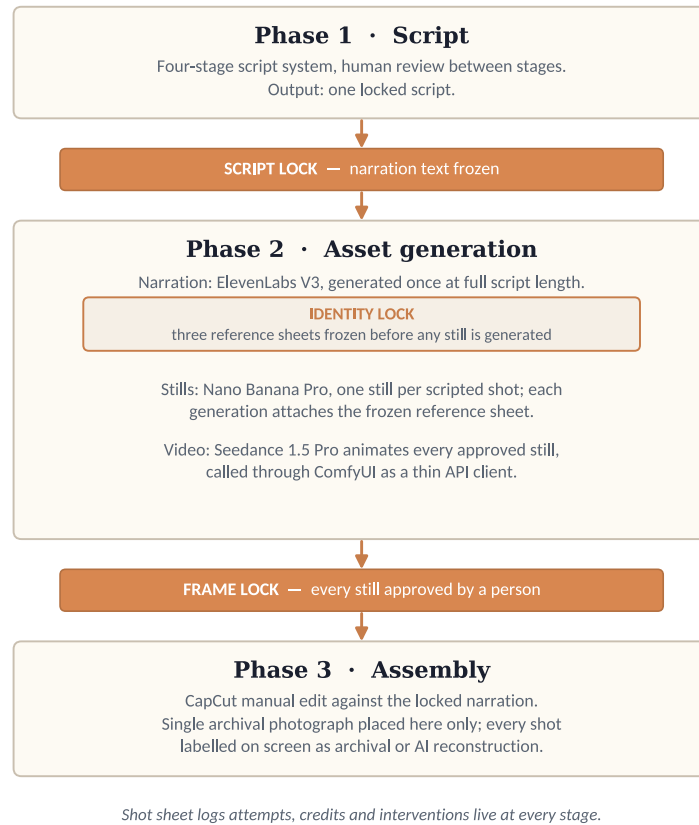


Figure 2. Pipeline architecture: three phases gated by the script lock, the identity lock and the frame lock.

One design principle repeats three times: freeze the approved asset so nothing changes as it moves down the pipeline. The script lock freezes narration before any audio is generated, because a late script change is the expensive kind of change, a cost pattern quantified for software four decades ago [17]. The identity lock freezes three character reference sheets before any scene is generated. The frozen file is the durable asset, because the hosted image model does not guarantee that a seed reproduces an image; so the saved sheet, instead of a seed value, is what actually preserves the identity. The frame lock requires every still to be approved by a person before it is animated. Naming that review as a gate did two jobs at once. It stopped visual drift from carrying through into video, and it turned every rejection into a logged data point for the robustness metric.

The single archival insert, a photograph of a public figure, enters at Phase 3 only. It is placed manually in the edit as an overlay on the closing clip, credited in the end titles, and recorded in the shot sheet like every other shot.

E. DATA COLLECTION AND PROCESSING

The shot sheet served as the evaluation instrument as well as the production manifest, so logging happened during generation, and the per-shot record holds credits, attempt count and any human intervention. The per-minute cost and regeneration figures follow from that record. Still-image latency was checked separately after the production run with two single-image workflows. One used one identity reference and one used no reference; each completed in 31 seconds.

Three evaluation dimensions were set in the proposal and retained for the final build. Efficiency: recorded visual generation spend and animation/edit time against an indicative vendor baseline. Quality: structured review against documentary standards of narration, coherence and pacing. Robustness: failure points and regeneration events over the run, in the sense used for system behaviour under real workload in [10]. The success criterion, also fixed in the proposal, was that the segment is produced inside the workflow with a clear advantage in one dimension or more.

F. ETHICAL AND LEGAL CONSIDERATIONS

The approved PG2 position was that the project handles no personal data, no biometrics and no identification of private individuals. The pipeline holds that position by construction, because no real person's face ever enters a generative model. Three tiers implement it. The lead character is a synthetic resemblance produced through the identity lock; the real defector's face is never used, neither generated nor shown. Public figures appear only as static archival photographs under Creative Commons licences, placed at assembly; the single insert in the artifact is a kremlin.ru photograph under CC BY 4.0 with a credit line in the end titles. Unidentified people are shown with faces concealed and are never described to a model. Disclosure is handled by a closing card, which states that all scenes are AI-generated reconstructions and credits the archival photograph to kremlin.ru under CC BY 4.0. The approach follows the Archival Producers Alliance guidance on generative material in documentary [18] and the transparency direction of Article 50 of the EU AI Act [19], which applies from August 2026 and is cited here as regulatory direction, since the UK is not bound by it. The Roadrunner controversy of 2021, in which a documentary used an undisclosed synthetic voice of Anthony Bourdain, showed the reputational cost of synthesis without disclosure [20].

Copyright needs stating in two halves, training and output. On training, litigation is live and courts have split by jurisdiction and facts. The Munich Regional Court ruled against OpenAI in November 2025, holding that lyrics memorised in model weights and reproduced in outputs infringe under German law; the ruling is under appeal [21]. In the United States the consolidated litigation against OpenAI remains in discovery, and district courts have divided on whether training is fair use. The UK government's March 2026 report on copyright and artificial intelligence surveys this position and confirms that commercial text and data mining in the UK still requires a licence [22]. This project trains no models and controls no training data; it consumes hosted services under the providers' terms. Google's Gemini API terms state that Google will not claim ownership over generated content [23].

On output, the UK answer is direct. Section 9(3) of the Copyright, Designs and Patents Act 1988 gives authorship of a computer-generated work to the person who undertakes the arrangements necessary for its creation, and for prompt-driven systems that is normally the person who wrote the prompts [24]. The government's 2026 report proposes removing section 9(3) in future legislation while keeping protection for AI-assisted works in which a human makes creative choices [22]. The provision remains in force. The documented arrangements support an authorship claim on either route: the script system, the prompts, the frozen reference sheets, the approvals at three gates and the assembly. The United States sets a stricter bar. In March 2026, the Supreme Court denied certiorari in *Thaler v. Perlmutter*, leaving the lower-court human-authorship ruling in place [25]. The documented human choices in selection, arrangement and editing are the elements a US claim could rest on.

V. IMPLEMENTATION AND DEVELOPMENT

A. THE PROTOTYPE

V1 was built in January against the D.B. Cooper story: 1 minute 15 seconds, produced quickly and roughly. Its purpose was structural. It proved that script, asset generation and assembly work as three separable phases with locked handoffs between them, and that a narration-led segment can be cut entirely from generated stills. Nothing from V1 was carried into the final artifact directly, but its three-phase structure was.

B. THE FULL BUILD ON SELF-HOSTED WEIGHTS

V2 ran from February to June and was the position presented at the interim viva. Images came from Flux.1 Dev [26], an open-weight diffusion model, running in ComfyUI [16] on RunPod [27] rented GPUs, an RTX 4090 at about \$0.50 an hour after 30 to 60 minutes of environment setup per session. Character consistency came from a LoRA [6] trained on a mannequin figure, driven by a trigger word placed first in every prompt, supported by negative blocklists and a pinned classifier-free guidance (CFG) value. Additional camera angles needed a second model: Qwen Image Edit [28] regenerated views from a master shot, two passes per angle.

The mannequin deserves explanation, because it was a style forced by a limit. The toolchain could not hold one photoreal face between shots; every attempt drifted. Training the LoRA on a featureless black matte mannequin removed the face entirely, and the artifact's early visual identity was born from that constraint. An unretouched V2 still can be seen in Figure 3 below.



Figure 3. Unretouched V2 output

On raw price V2 was strong: about 400 stills for \$0.85 of GPU time. The system was retired in week 24 over four structural faults. There was no stable face without a training step. Every extra angle cost two model passes. Every session began with setup before any image existed. And every prompt needed babysitting: trigger placement, negative lists, CFG pinning. Working sessions had a fixed rhythm: rent the GPU, rebuild the environment, load the LoRA, then generate against a prompt template whose trigger placement, blocklists and CFG value could not safely change. The rhythm worked, and it put half an hour of ceremony in front of every hour of output. By the interim viva the pipeline worked, but its visual style existed mainly as a workaround for the face problem.

C. THE V3 PIVOT

The pivot was planned, tested and executed in early July. Nano Banana Pro [14] accepts reference images as an input: attach the frozen sheet, describe the shot in plain language, and the face comes back right, with no training step. V2 had to bake identity into LoRA weights to get the same property. Table 3 sets out the swap stage by stage.

Stage	V2	V3
Images	Flux.1 Dev + mannequin LoRA	Nano Banana Pro
Consistency	LoRA weights, trigger word	Frozen reference sheets
Angles	Qwen two-stage edit	Plain language + master shot
Compute	RunPod rented GPUs	Hosted API, credit billed
Look	Black matte mannequin	Photoreal synthetic human
Video	Kling only	Seedance 1.5 Pro, by test

Table 3. The V2 to V3 swap, stage by stage.

The identity lock was the central design decision of the pivot. Three character reference sheets were generated, reviewed and frozen before any scene work began. The hosted model does not promise that a seed will bring back the same image, so a seed value cannot be relied on to reproduce an image. The saved file is the only durable copy of the identity. Every scene generation attaches the sheet, and the lead's identity held over the full run with zero retraining. Figure 4 below shows the frozen reference sheet beside an in-scene frame, both unretouched.



Figure 4. Frozen identity sheet next to the same character in scene.

The proposal claimed the pipeline's stages were interchangeable. The swap proved it. The visual branch was rebuilt in days, the script and narration phases did not change by a line, and the modularity claim moved from design intention to demonstrated property. Angles simplified in the same move: where V2 ran a two-pass edit model, V3 attaches the master shot beside the identity sheet and asks for the new view in a sentence.

D. VIDEO SELECTION AND THE PRODUCTION RUN

Video was decided by a one-scene head-to-head on the same approved frame: Seedance 1.5 Pro against Seedance 2.0, measured on time, cost, quality and intervention count. The result was equal fidelity at a fifth of the cost, so Seedance 1.5 Pro [15] animated every clip in the artifact at 54.9 credits per five-second generation. Kling [29], the V2 video model, was held as the fallback and never called. The rule, fixed in advance, was to test one scene, commit to the result and not switch back. No switches were needed during the run.

Generation ran through four batch workflows in ComfyUI, acting purely as an API client: stills with references, stills without references, essential motion and subtle motion. The video node chain is three nodes long: LoadImage, ByteDanceImageToVideoNode, SaveVideo. The run finished on 9 July. Stills went first, in batches against the shot list, and each batch passed the frame lock before its approved frames queued for animation through the essential-motion workflow; the subtle-motion workflow was held as a fallback and never taken. Logging ran live, each asset landing in the shot sheet with its attempts and credits as it arrived. Narration had been generated once, months earlier, at full script length for about \$6.50, and was never touched again; the script lock made that possible.

E. LESSON APPLIED: MERGED BEATS SPLIT

The prompt system itself went through the project's sharpest design lesson. The early pipeline split editorial decisions over three components: a scene breakdown generator, an image prompt writer and an animation prompt writer. Splitting decisions about a single shot over three components meant the same editorial error could be introduced three times, so the components were merged and each decision about a shot is now made in one place. The breakdown stage was retired, the animation pass was absorbed, and the reference sheets were kept inside a single merged Prompt Writer, version 9 by the end. The per-shot output is the full record the batch workflows consume: the editorial description, the image prompt, the motion instruction and the reference sheet the generation must attach. The merged writer's shot list is the manifest the rest of the pipeline runs on, Appendix C shows one complete per-shot record, the Prompt Writer V9 output for shot 0019, alongside the batch workflows that use it.

VI. TESTING AND EVALUATION

A. THE ARTIFACT AND THE HEADLINE NUMBERS

The artifact is 'The Pilot Who Flew Home', a 3 minute 18 second documentary segment. Appendix F holds the access link to the finished cut. The production run generated 57 stills and logged one archival photograph, 58 shot-sheet rows in all. Of the 57 stills, 55 were animated and two, shots 36 and 37, were held static with a Ken Burns move applied in CapCut. The shot sheet had planned 32 static rows against 25 motion rows, and the plan did not survive contact with production: once a still is frame-locked, animating it costs a flat 54.9 credits, so 30 of the planned statics were animated anyway.

The final timeline holds 44 shots, 42 animated clips and the two Ken Burns statics, with the archival photograph placed as an overlay on the closing clip. The edit dropped 13 of the 55 clips, a shooting ratio of 1.3 clips generated per clip used, and retimed several of the clips it kept. Every dropped clip had passed the frame lock; the cuts were pacing decisions made against the narration, and at 54.9 credits per clip the over-coverage was cheap insurance for the edit. The \$9.56 visual generation spend figure includes the dropped clips because it was computed from total credits consumed. Table 4 gives the headline results, all derived from the shot sheet.

Measure	Result
Finished runtime	3 min 18 s
Visual generation spend per finished minute	\$9.56
Recorded animation and edit time per finished minute	73 min
Regeneration events per finished minute	3.0
Stills passed first take	55 of 57
Clips passed first take	51 of 55
Clips used in the final cut	42 of 55
Identity retraining events	0

Table 4. Headline results, derived from the shot sheet.

Definitions are needed for three measures. The \$9.56 figure is the visual branch's full generation spend, including tests and retries, on a conservative balance basis: roughly 6,660 credits consumed, about \$31.6 at \$0.004739 per credit, divided by 3.3 finished minutes. Narration is excluded because one \$6.50 generation at full script length serves every future segment of the episode. The 73-minute figure covers animation and CapCut assembly. The full stills production session was not timed. Two later single-image checks each completed in 31 seconds, one with one identity reference and one with no reference. They measure service latency under the tested conditions. The production workflows launched image-generation nodes concurrently, so the latency was shared by the images in each batch. Prompt writing, queueing, approvals and batch handling remain excluded. The intervention measure counts regeneration events only; prompt writing, approvals and editing sit outside it. A total production-cost comparison was unavailable because operator labour and total still-image production time were not logged. Set against Table 2, six requirements were met in full: F1, F2, F5, F6, N1 and N2. F3 was met for the 57 generated shots; the archival row enters at assembly by design. F4 was partly met because shots 36 and 37 passed the frame lock and were held static with a Ken Burns move. N3 was supported by one demonstrated replacement: the V2 visual branch was replaced without changing the locked script or narration stages. N4 was partly met because the balance basis captures total spend, while about 1,550 credits of test and retry consumption remain unitemised.

B. EFFICIENCY

Money. Published vendor estimates place conventional documentary production between \$1,000 and \$10,000 per finished minute [1], [2]. They provide an indicative baseline because they include crew, equipment, travel and licensing. The pipeline's visual generation spend was \$9.56 per finished minute, two orders of magnitude below the lowest estimate. A five-second clip costs a flat 54.9 credits, about \$0.26, so a shot's price is known before generation. The cost scopes differ. The pipeline figure excludes operator labour and unlogged still-image production time. Figure 5 plots the figures on that stated basis.

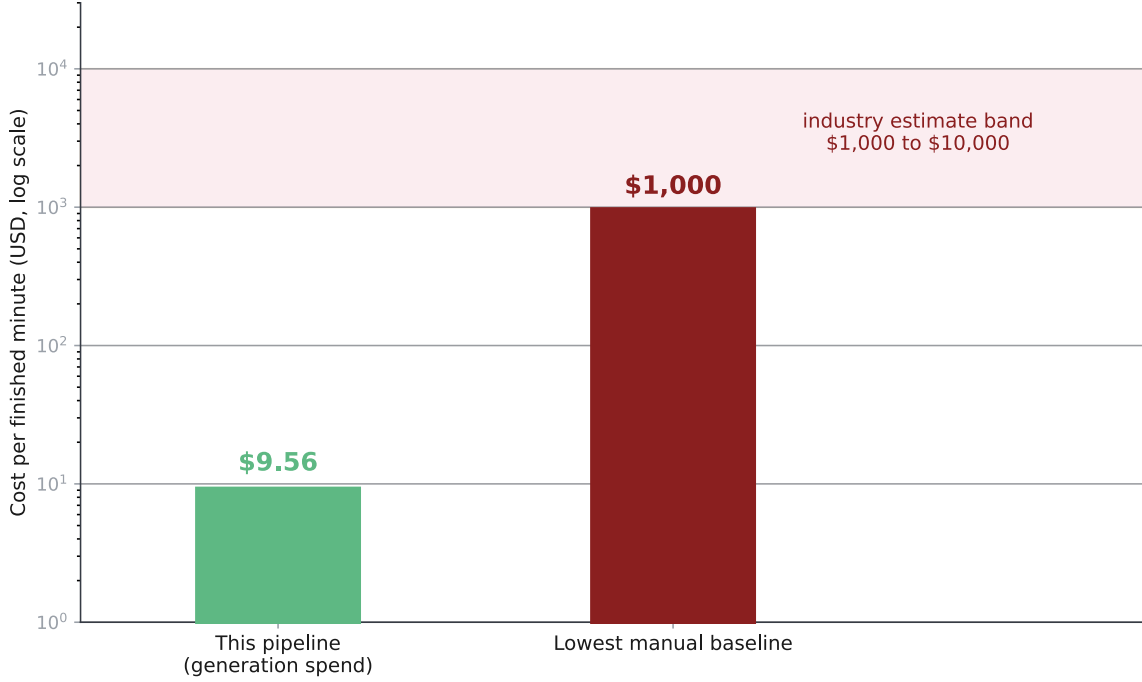


Figure 5. Visual generation spend per finished minute against the indicative vendor baseline band, on a logarithmic scale.

Time. Recorded animation and edit work took 73 minutes per finished minute, about four hours for the 3:18 cut, against a conventional post-production estimate of at least one editing day, 8 hours or more, per finished minute [2]. Within the recorded scope, the animation-and-edit stage was roughly six to seven times shorter, and the pipeline required no shoot. Two later latency checks each completed in 31 seconds, one with an identity reference and one without. Production stills were submitted through batch workflows whose generation nodes ran concurrently, so the 31-second latency applied to a submitted batch instead of each individual image. Total still-production time remains unrecorded because batch submissions, retries, queueing, prompt writing and approvals were not timed. The 73-minute measure remains limited to animation and CapCut assembly.

Episode projection. On a linear assumption, a 45-minute episode projects to about \$430 in visual generation spend, about 55 hours of recorded animation and edit work, and about 135 regeneration events. The identity sheets and the one-scene model test are fixed costs that a full episode would spread over a longer runtime. The linear assumption is untested above 3.3 minutes; the projection states scale, and the limitations section states its weakness.

C. QUALITY

Quality was assessed by structured review against the proposal's documentary standards, narration, coherence and pacing, and by screening at the final viva demonstration. The review was a structured self-assessment by the author; no external panel scored the artifact. Narration was judged on clarity and register, coherence on story logic and visual continuity, pacing on shot rhythm against the narration. Narration held a single consistent voice from one full-length generation. Visual coherence rests on the identity lock: the lead's identity held over the full run with zero retraining, and no shot was rejected for identity drift. Pacing settled at an average shot length of 4.5

seconds over the 198-second, 44-shot cut, established manually in the edit. Of the three dimensions, quality carries the least numerical evidence, and this is acknowledged as a limitation; the structured review and the screened demo are its record.

D. ROBUSTNESS

Robustness was measured as failure points and regeneration events over the run. 55 of 57 stills and 51 of 55 clips passed first take; every rejection was regenerated inside the credit log. Regeneration events averaged 3.0 per finished minute, 10 events over the run: five on the stills pass, concentrated in shots 0010 and 0025, and five on the video pass, shots 0001, 0002, 0003 and 0040. Every event is recorded against its shot in Appendix A. The run recorded zero identity drift, zero retraining and zero model switches after the one-scene commitment.

One data-integrity finding belongs on the record. A pre-submission audit found that shot 0002 logged one attempt while credit records show two generations, so the shot sheet's logged total understated spend by 27.0 credits. Appendix A carries the corrected sheet. Appendix B reconciles the full accounts: 8,440 credits deposited over four top-ups, about 1,780 remaining, roughly 6,660 consumed on a balance basis, of which 4,937.0 is itemised on the shot sheet and 169.8 is identified off-sheet spend on the three identity sheets and the model test. The remainder is unitemised test and retry consumption. The visual generation spend figure uses the balance basis, so the unitemised spend remains inside the reported total.

E. LIMITATIONS

Six limitations close this evaluation. Editing still dominates the recorded human hours; the manual CapCut work is the pipeline's largest recorded time cost and its next automation target. The 73-minute figure excludes the stills production session. The two 31-second checks give a service-latency sample, while the concurrent batch design prevents that figure from being added once per image. Total operator time for prompt writing, queueing, approvals and batch handling was not logged. Tool hops remain manual between ComfyUI and CapCut; the Conclusions and Recommendations section proposes an orchestration screen to close them. Seed non-determinism in the hosted image model makes the frozen file the only durable identity asset, which depends on careful file management. The provider can update or retire a hosted model at any time; V3 traded control of local weights for capability. The sixth constraint bounds the whole evaluation: every number comes from one 3.3-minute segment, and the episode projection assumes linear scaling that a longer run has not tested.

VII. CONCLUSIONS AND RECOMMENDATIONS

A. CONCLUSIONS

Four findings answer the research question. A locked script became a finished segment with \$9.56 visual generation spend, 73 minutes of recorded animation and edit work and 3.0 regeneration events per finished minute. The visual branch was replaced without changing the script or narration phases, which supports modularity in practice. The three locks provide a reusable design pattern: freeze the approved asset upstream, consume it downstream and log each rejection at the gate. The success criterion was met on efficiency against an indicative vendor baseline, with robustness measured and quality reviewed from the production log. The final build met its operational objectives. Episode-length production and full end-to-end automation from the original proposal remain untested.

B. RECOMMENDATIONS AND FUTURE WORK

The next build is an orchestration interface: one screen holding every stage, gate and credit balance, with the existing batch runners behind it. Figure 6 below shows the proposed screen as a wireframe. It is designed but deliberately unbuilt, so that nothing unevaluated is claimed as part of the artifact.

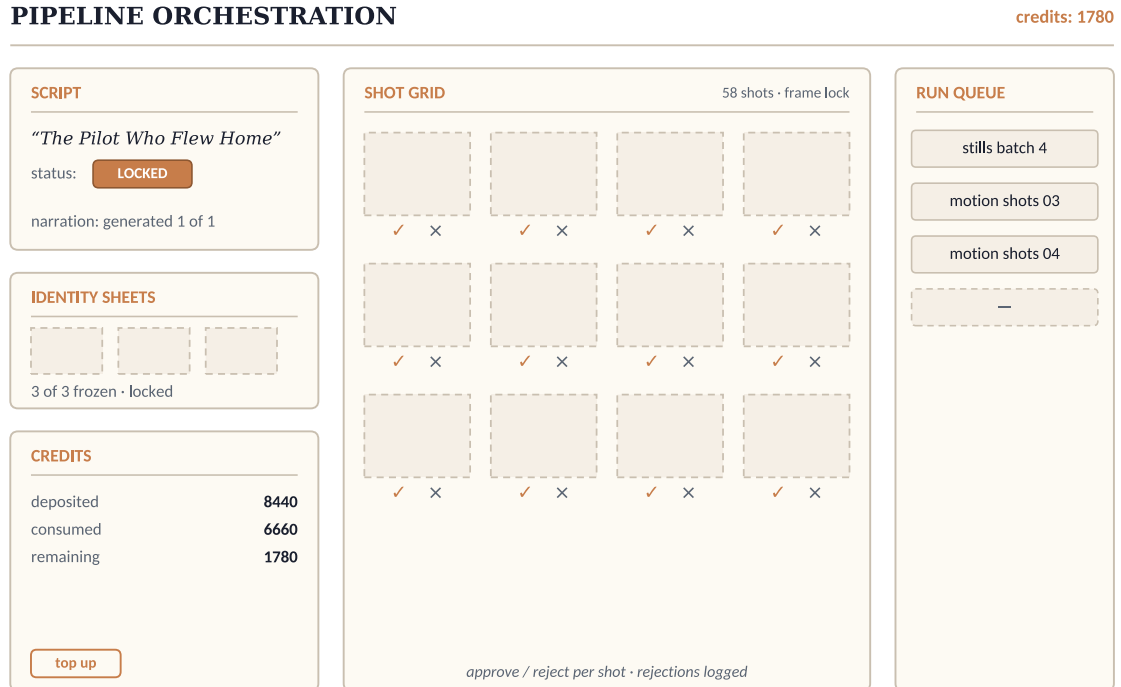


Figure 6. The proposed orchestration screen, designed and deliberately unbuilt.

Beyond the interface: scripted handoffs end to end, so the manual tool hops disappear; automated quality checks that flag identity drift before human review, so the frame lock spends attention only where it is needed; a full-episode production run to test the linear projection; and Seedance 2.5 once it is publicly available. The architecture also generalises past this module. As a service, the same three gates become customer approval points: bring a story, approve at three gates, leave with an episode, at a price without a studio behind it.

VIII. REFLECTION

The most useful lesson I took from the module concerned measurement: it has to be designed in from the start, because it cannot be added afterwards. Because I built the shot sheet as the evaluation log from the first generation, the per-minute figures in Testing and Evaluation could be calculated by simple division at the end, with nothing needing to be reconstructed afterwards. The frame lock taught me the same lesson from the other side: it began as an informal look at each still, and its rejections only became usable data once I formalised it as a gate.

The hardest decision I made was retiring V2 in week 24. Five months of working pipeline went with it, and the choice felt expensive until I wrote the comparison down. Seeing the four faults listed side by side made it much easier to accept. Committing to Seedance 1.5 Pro after a single one-scene test took the same discipline at smaller scale, and the harder part was sticking to that commitment for the rest of the run. I managed the project over 27 weeks, from script lock at week 5 through the interim viva at week 18 to the pivot and finished artifact in week 24, and the schedule only had to change once, at the pivot, and the rest of the plan held. The final Gantt chart in Appendix E records the plan and the bend.

On a second run I would change two things. I would name the frame lock as a gate from day one, and I would test hosted alternatives one full cycle earlier, since the capability that ended V2 existed before V2 was finished. Building on hosted models means the underlying services change from month to month. Modular design, small comparative tests and firm commitment afterwards turned out to be the best protection against that.

IX. REFERENCES

- [1] Beverly Boy Productions. "What is the cost of documentary film production?" Beverly Boy Productions. Accessed: 17 July 2026. [Online]. Available: <https://beverlyboy.com/video-production/what-is-the-cost-of-documentary-film-production/>
- [2] Wind & Sky Productions. "How much does a documentary cost to make?" Wind & Sky Productions. Accessed: 17 July 2026. [Online]. Available: <https://windsky.com.au/how-much-does-a-documentary-cost/>
- [3] L. Yang *et al.*, "Diffusion models: A comprehensive survey of methods and applications," *ACM Comput. Surv.*, vol. 56, no. 4, pp. 1–39, 2024, doi: 10.1145/3626235.
- [4] H. Barakat, O. Turk and C. Demiroglu, "Deep learning-based expressive speech synthesis: a systematic review of approaches, challenges, and resources," *EURASIP J. Audio Speech Music Process.*, vol. 2024, no. 1, Art. no. 11, 2024, doi: 10.1186/s13636-024-00329-7.
- [5] W. X. Zhao *et al.*, "A survey of large language models," arXiv:2303.18223, 2023, doi: 10.48550/arXiv.2303.18223.
- [6] E. J. Hu *et al.*, "LoRA: Low-rank adaptation of large language models," arXiv:2106.09685, 2021, doi: 10.48550/arXiv.2106.09685.
- [7] H. Ye, J. Zhang, S. Liu, X. Han and W. Yang, "IP-Adapter: Text compatible image prompt adapter for text-to-image diffusion models," arXiv:2308.06721, 2023, doi: 10.48550/arXiv.2308.06721.
- [8] A. Blattmann *et al.*, "Stable video diffusion: Scaling latent video diffusion models to large datasets," arXiv:2311.15127, 2023, doi: 10.48550/arXiv.2311.15127.
- [9] C. Hight, "Deepfakes and documentary practice in an age of misinformation," *Continuum*, vol. 36, no. 3, pp. 393–410, 2022, doi: 10.1080/10304312.2021.2003756.
- [10] A. Bandi, P. V. S. R. Adapa and Y. E. V. P. K. Kuchi, "The Power of Generative AI: A Review of Requirements, Models, Input–Output Formats, Evaluation Metrics, and Challenges," *Future Internet*, vol. 15, no. 8, Art. no. 260, 2023, doi: 10.3390/fi15080260.
- [11] A. R. Hevner, S. T. March, J. Park and S. Ram, "Design Science in Information Systems Research," *MIS Q.*, vol. 28, no. 1, pp. 75–105, 2004, doi: 10.2307/25148625.
- [12] K. Peffers, T. Tuunanen, M. A. Rothenberger and S. Chatterjee, "A Design Science Research Methodology for Information Systems Research," *J. Manage. Inf. Syst.*, vol. 24, no. 3, pp. 45–77, 2007, doi: 10.2753/MIS0742-1222240302.
- [13] ElevenLabs. "ElevenLabs Text to Speech." elevenlabs.io. Accessed: 21 March 2026. [Online]. Available: <https://elevenlabs.io>
- [14] Google DeepMind. "Gemini 3 Pro Image (Nano Banana Pro)." Google DeepMind. Accessed: 17 July 2026. [Online]. Available: <https://deepmind.google/models/gemini-image/pro/>
- [15] ByteDance Seed Team. "Seedance 1.5 pro." ByteDance Seed. Accessed: 17 July 2026. [Online]. Available: https://seed.bytedance.com/en/seedance1_5_pro
- [16] Comfy-Org. "ComfyUI." GitHub. Accessed: 17 July 2026. [Online]. Available: <https://github.com/Comfy-Org/ComfyUI>
- [17] B. W. Boehm, "Software Engineering Economics," *IEEE Trans. Softw. Eng.*, vol. SE-10, no. 1, pp. 4–21, 1984, doi: 10.1109/TSE.1984.5010193.

- [18] Archival Producers Alliance. "Best practices guidelines for use of generative AI in documentaries." Archival Producers Alliance. Accessed: 17 July 2026. [Online]. Available: <https://www.archivalproducersalliance.com/genai-guidelines>
- [19] European Parliament and Council of the European Union, "Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence (Artificial Intelligence Act)," Luxembourg, 2024. Accessed: 17 July 2026. [Online]. Available: <https://eur-lex.europa.eu/eli/reg/2024/1689/oj>
- [20] H. Rosner, "The ethics of a deepfake Anthony Bourdain voice," The New Yorker, 2021. Accessed: 18 July 2026. [Online]. Available: <https://www.newyorker.com/culture/annals-of-gastronomy/the-ethics-of-a-deepfake-anthony-bourdain-voice>
- [21] Bird & Bird. "Landmark ruling of the Munich Regional Court (GEMA v OpenAI) on copyright and AI training." Bird & Bird. Accessed: 17 July 2026. [Online]. Available: [https://www.twobirds.com/en/insights/2025/landmark-ruling-of-the-munich-regional-court-\(gema-v-openai\)-on-copyright-and-ai-training](https://www.twobirds.com/en/insights/2025/landmark-ruling-of-the-munich-regional-court-(gema-v-openai)-on-copyright-and-ai-training)
- [22] Department for Science, Innovation and Technology, Department for Culture, Media and Sport, and Intellectual Property Office, "Report and Impact Assessment on Copyright and Artificial Intelligence," London, United Kingdom, 2026. Accessed: 17 July 2026. [Online]. Available: <https://www.gov.uk/government/publications/report-and-impact-assessment-on-copyright-and-artificial-intelligence>
- [23] Google. "Gemini API additional terms of service." Google AI for Developers. Accessed: 17 July 2026. [Online]. Available: <https://ai.google.dev/gemini-api/terms>
- [24] Copyright, Designs and Patents Act 1988, c. 48, s. 9(3). UK Public General Acts, 1988. Accessed: 18 July 2026. [Online]. Available: <https://www.legislation.gov.uk/ukpga/1988/48/section/9>
- [25] Supreme Court of the United States. "Thaler v. Perlmutter, No. 25-449, docket." supremecourt.gov. Accessed: 18 July 2026. [Online]. Available: <https://www.supremecourt.gov/docket/docketfiles/html/public/25-449.html>
- [26] Black Forest Labs. "FLUX.1 Dev." huggingface.co. Accessed: 21 March 2026. [Online]. Available: <https://huggingface.co/black-forest-labs/FLUX.1-dev>
- [27] RunPod. "RunPod." runpod.io. Accessed: 21 March 2026. [Online]. Available: <https://www.runpod.io>
- [28] Qwen Team. "Qwen-Image-Edit." Hugging Face. Accessed: 17 July 2026. [Online]. Available: <https://huggingface.co/Qwen/Qwen-Image-Edit>
- [29] Kuaishou Technology. "Kling AI." klingai.com. Accessed: 21 March 2026. [Online]. Available: <https://klingai.com>

X. APPENDICES

A. THE FULL SHOT SHEET

The production run covers 58 shot-sheet rows, 57 AI-generated stills and one archival photograph, shot 0058, placed at assembly as an overlay on the closing clip. Figure A.1 is the contact sheet of the final approved frame for every shot. Table A.1 is the shot sheet as logged during generation, one row per shot, with attempt counts and credit spend for the stills pass and the video pass. Credits were billed per generation: 28.3 per still, 54.9 per clip with audio, 27.0 per clip without audio. Shots 0036 and 0037 were held static and carry no video spend; the archival shot carries no generation spend at all. The table is the production record. The final cut uses 42 of the 55 animated clips plus the two statics; the edit decisions behind that selection are reported in Testing and Evaluation.

One row is shown corrected. Credit records for shot 0002 held two video generations, 54.9 with audio and 27.0 without, against a single logged attempt, so the row now reads two attempts and 81.9 credits. Appendix B carries the full credit reconciliation, including this correction.

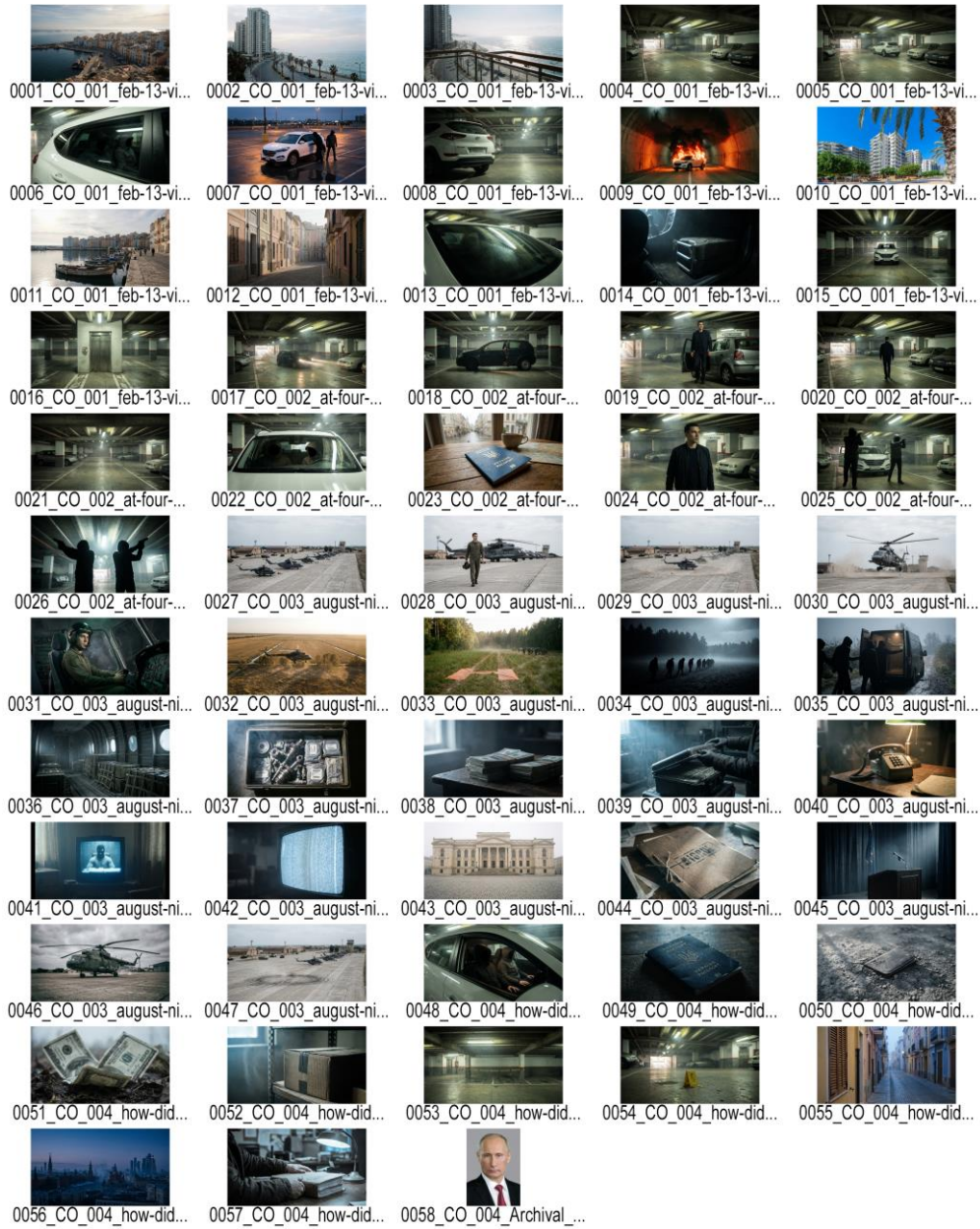


Figure A.1. Contact sheet of the final approved frame for each of the 58 shots, in shot order.

Shot ID	Planned type	Still att.	Still credits	Video att.	Video credits	Notes
0001_CO_001_feb-13-villajoyosa-spain_SC01	STILL	1	28.3	3	164.7	
0002_CO_001_feb-13-villajoyosa-spain_SC02	STILL	1	28.3	2	81.9	Corrected pre-submission: credit records show two generations (54.9 with audio, 27.0 without) against one logged attempt; attempts corrected 1 to 2. See Appendix B.
0003_CO_001_feb-13-villajoyosa-spain_SC03	STILL	1	28.3	2	81.9	2nd attempt without audio, it's cheaper, only 27 credits
0004_CO_001_feb-13-villajoyosa-spain_SC04	STILL	1	28.3	1	54.9	
0005_CO_001_feb-13-villajoyosa-spain_SC05	MOTION_SUBTLE	1	28.3	1	54.9	
0006_CO_001_feb-13-villajoyosa-spain_SC06	MOTION_SUBTLE	1	28.3	1	54.9	
0007_CO_001_feb-13-villajoyosa-spain_SC07	STILL	1	28.3	1	54.9	
0008_CO_001_feb-13-villajoyosa-spain_SC08	STILL	1	28.3	1	54.9	
0009_CO_001_feb-13-villajoyosa-spain_SC09	MOTION_ESSENTIAL	1	28.3	1	54.9	
0010_CO_001_feb-13-villajoyosa-spain_SC10	STILL	2	56.6	1	27	without audio
0011_CO_001_feb-13-villajoyosa-spain_SC11	STILL	1	28.3	1	27	without audio
0012_CO_001_feb-13-villajoyosa-spain_SC12	STILL	1	28.3	1	54.9	
0013_CO_001_feb-13-villajoyosa-spain_SC13	STILL	1	28.3	1	54.9	
0014_CO_001_feb-13-villajoyosa-spain_SC14	STILL	1	28.3	1	54.9	
0015_CO_001_feb-13-villajoyosa-spain_SC15	MOTION_SUBTLE	1	28.3	1	54.9	

Shot ID	Planned type	Still att.	Still credits	Video att.	Video credits	Notes
0016_CO_001_feb-13-villajoyosa-spain_SC16	STILL	1	28.3	1	54.9	
0017_CO_002_at-four-twenty-pm_SC01	MOTION_ESSENTIAL	1	28.3	1	54.9	
0018_CO_002_at-four-twenty-pm_SC02	MOTION_SUBTLE	1	28.3	1	54.9	
0019_CO_002_at-four-twenty-pm_SC03	MOTION_ESSENTIAL	1	28.3	1	54.9	
0020_CO_002_at-four-twenty-pm_SC04	MOTION_ESSENTIAL	1	28.3	1	54.9	
0021_CO_002_at-four-twenty-pm_SC05	STILL	1	28.3	1	54.9	
0022_CO_002_at-four-twenty-pm_SC06	MOTION_SUBTLE	1	28.3	1	54.9	
0023_CO_002_at-four-twenty-pm_SC07	STILL	1	28.3	1	54.9	
0024_CO_002_at-four-twenty-pm_SC08	MOTION_ESSENTIAL	1	28.3	1	54.9	
0025_CO_002_at-four-twenty-pm_SC09	MOTION_ESSENTIAL	5	141.5	1	54.9	
0026_CO_002_at-four-twenty-pm_SC10	STILL	1	28.3	1	54.9	
0027_CO_003_august-ninth-helicopter_SC01	STILL	1	28.3	1	54.9	
0028_CO_003_august-ninth-helicopter_SC02	MOTION_ESSENTIAL	1	28.3	1	54.9	
0029_CO_003_august-ninth-helicopter_SC03	MOTION_SUBTLE	1	28.3	1	54.9	
0030_CO_003_august-ninth-helicopter_SC04	MOTION_ESSENTIAL	1	28.3	1	54.9	
0031_CO_003_august-ninth-helicopter_SC05	MOTION_SUBTLE	1	28.3	1	54.9	
0032_CO_003_august-ninth-helicopter_SC06	MOTION_ESSENTIAL	1	28.3	1	54.9	
0033_CO_003_august-ninth-helicopter_SC07	STILL	1	28.3	1	54.9	
0034_CO_003_august-ninth-helicopter_SC08	MOTION_SUBTLE	1	28.3	1	54.9	
0035_CO_003_august-ninth-helicopter_SC09	MOTION_SUBTLE	1	28.3	1	54.9	

Shot ID	Planned type	Still att.	Still credits	Video att.	Video credits	Notes
0036_CO_003_august-ninth-helicopter_SC10	STILL	1	28.3	0	0	
0037_CO_003_august-ninth-helicopter_SC11	STILL	1	28.3	0	0	
0038_CO_003_august-ninth-helicopter_SC12	STILL	1	28.3	1	54.9	
0039_CO_003_august-ninth-helicopter_SC13	MOTION_SUBTLE	1	28.3	1	54.9	
0040_CO_003_august-ninth-helicopter_SC14	MOTION_SUBTLE	1	28.3	2	109.8	
0041_CO_003_august-ninth-helicopter_SC15	MOTION_SUBTLE	1	28.3	1	54.9	
0042_CO_003_august-ninth-helicopter_SC16	STILL	1	28.3	1	54.9	
0043_CO_003_august-ninth-helicopter_SC17	STILL	1	28.3	1	54.9	
0044_CO_003_august-ninth-helicopter_SC18	STILL	1	28.3	1	54.9	
0045_CO_003_august-ninth-helicopter_SC19	STILL	1	28.3	1	54.9	
0046_CO_003_august-ninth-helicopter_SC20	MOTION_SUBTLE	1	28.3	1	54.9	
0047_CO_003_august-ninth-helicopter_SC21	STILL	1	28.3	1	54.9	
0048_CO_004_how-did-two-men_SC01	MOTION_SUBTLE	1	28.3	1	54.9	
0049_CO_004_how-did-two-men_SC02	STILL	1	28.3	1	54.9	
0050_CO_004_how-did-two-men_SC03	STILL	1	28.3	1	54.9	
0051_CO_004_how-did-two-men_SC04	STILL	1	28.3	1	54.9	
0052_CO_004_how-did-two-men_SC05	STILL	1	28.3	1	54.9	
0053_CO_004_how-did-two-men_SC06	MOTION_SUBTLE	1	28.3	1	54.9	
0054_CO_004_how-did-two-men_SC07	STILL	1	28.3	1	54.9	
0055_CO_004_how-did-two-men_SC08	STILL	1	28.3	1	54.9	

Shot ID	Planned type	Still att.	Still credits	Video att.	Video credits	Notes
0056_CO_004_how-did-two-men_SC09	STILL	1	28.3	1	54.9	
0057_CO_004_how-did-two-men_SC10	MOTION_SUBTLE	1	28.3	1	54.9	
0058_CO_004_how-did-two-men_SC11	ARCHIVAL	—	0	—	0	ARCHIVAL. Source: Presidential Photo bank, kremlin.ru, CC BY 4.0 (http://static.kremlin.ru/media/events/press-photos/orig/41d3e9385e34ebc0e3ba.jpeg?kremlin-ru_photo-1.jpeg); no generation. Placement: over the closing line "what it actually costs to betray Vladimir Putin," held on the final beat of the cold open. Apply slow Ken Burns push-in in CapCut.
Totals		62	1,754.6	60	3,182.4	Grand total: 4,937.0 credits itemised on-sheet.

Table A.1. The full shot sheet, 58 rows, with attempts and credits per shot as logged during generation.

B. CREDIT ACCOUNTING RECONCILIATION

Credits were bought in four \$10 top-ups at the platform rate of 2,110 credits per \$10, a deposit of 8,440.0 credits and \$40 in total. The account balance at the close of production read approximately 1,780 credits, which puts consumption at approximately 6,660. Table B.1 sets that balance-derived figure against the itemised records.

Item	Credits
Deposited: four \$10 top-ups at 2,110 credits per \$10	8,440.0
Balance remaining at close of production	~1,780
Consumed, derived from balance (deposited minus remaining)	~6,660
Itemised on-sheet in Table A.1 (stills 1,754.6; video 3,182.4)	4,937.0
Itemised off-sheet: three frozen identity sheets and the one-scene video model test	169.8
Total itemised	5,106.8
Unitemised remainder (consumed minus total itemised)	~1,553

Table B.1. Credit reconciliation between the account balance and the itemised production records.

The unitemised remainder is spend the credit records show but no production log claims: generations run outside the logged shot-sheet workflow. It is reported as a gap, since reconstructing it after the fact would produce numbers with no record behind them.

One correction crossed from this reconciliation into Table A.1. The credit records for shot 0002 hold two video generations, 54.9 with audio and 27.0 without, against a single logged attempt. The row was corrected before submission to two attempts and 81.9 credits; the on-sheet total of 4,937.0 already carries the correction.

C. PROMPT SYSTEM

Prompt Writer V9 wrote every image prompt and every video prompt on the shot sheet in one editorial pass over the locked script. Table A.1 drops the two prompt columns for page width, so the full record for one representative shot is reproduced here as Table C.1. Shot 0019 shows Kuzminov stepping out of the parked car in the garage; the row draws on the frozen identity sheet in Appendix D and the garage environment master at the same time, which makes it a fair sample of how the reference plan, the image prompt and the video prompt sit together on a single row.

Field	Content
ID	0019_CO_002_at-four-twenty-pm_SC03
Section	{CO_NAR_at-four-twenty-pm}
Type	MOTION_ESSENTIAL, 5 seconds
References	1: SHEET_KUZMINOV, 2: ENV_GARAGE_CO
Image prompt	The man from the first reference image, wearing a dark civilian jacket, plain crew-neck shirt and dark trousers, stepping out of the parked car and standing up in the environment from the second reference image, seen in a medium shot from a low angle, weak fluorescent light from above, faint exhaust haze, 35mm lens, shallow depth of field, cinematic documentary still, natural film grain, desaturated midtones, high contrast, photorealistic
Video prompt	The man rises out of the car and pushes the door shut behind him, settling onto his feet. The camera holds a slow low push-in. His face, clothing and the garage remain exactly as in the source frame.
Production record	Still: 1 attempt, 28.3 credits. Video: 1 attempt, 54.9 credits.

Table C.1. The complete Prompt Writer V9 record for shot 0019, as delivered in shot_sheet_chunk_01 and filled during production.

Figures C.1 to C.3 show three of the four batch workflows that consumed these prompts; the subtle-motion fallback was never taken and is not pictured.

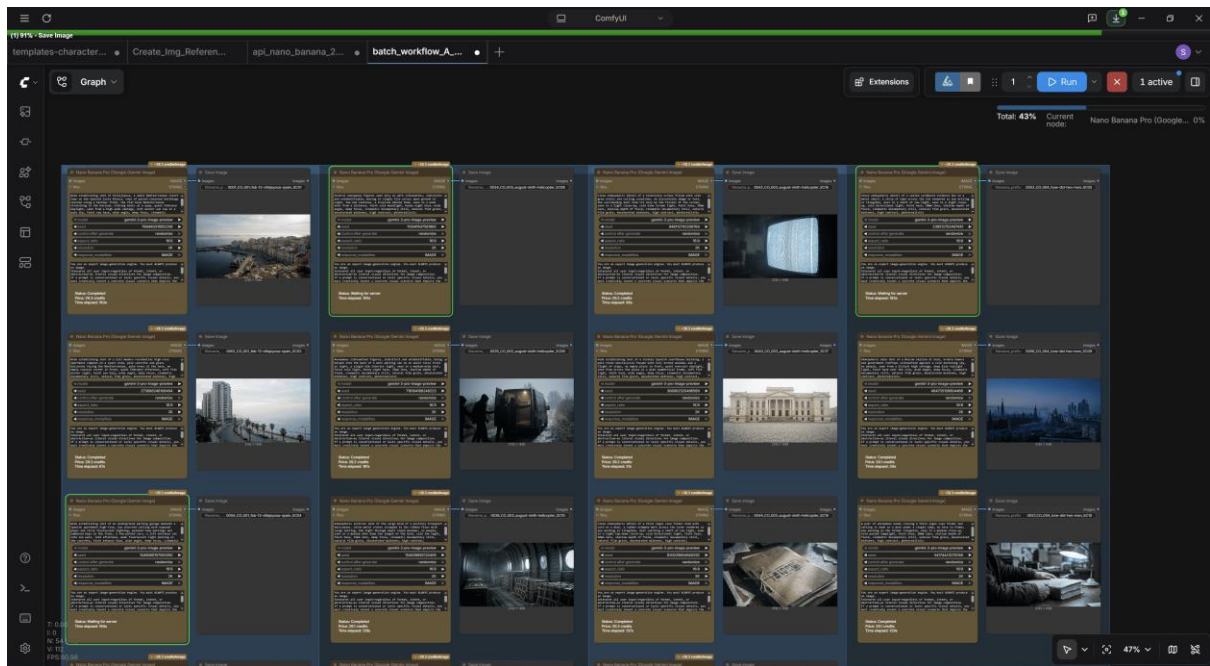


Figure C.1. ComfyUI batch workflow for the stills pass without references, sending per-shot prompts to Nano Banana Pro through the hosted API.

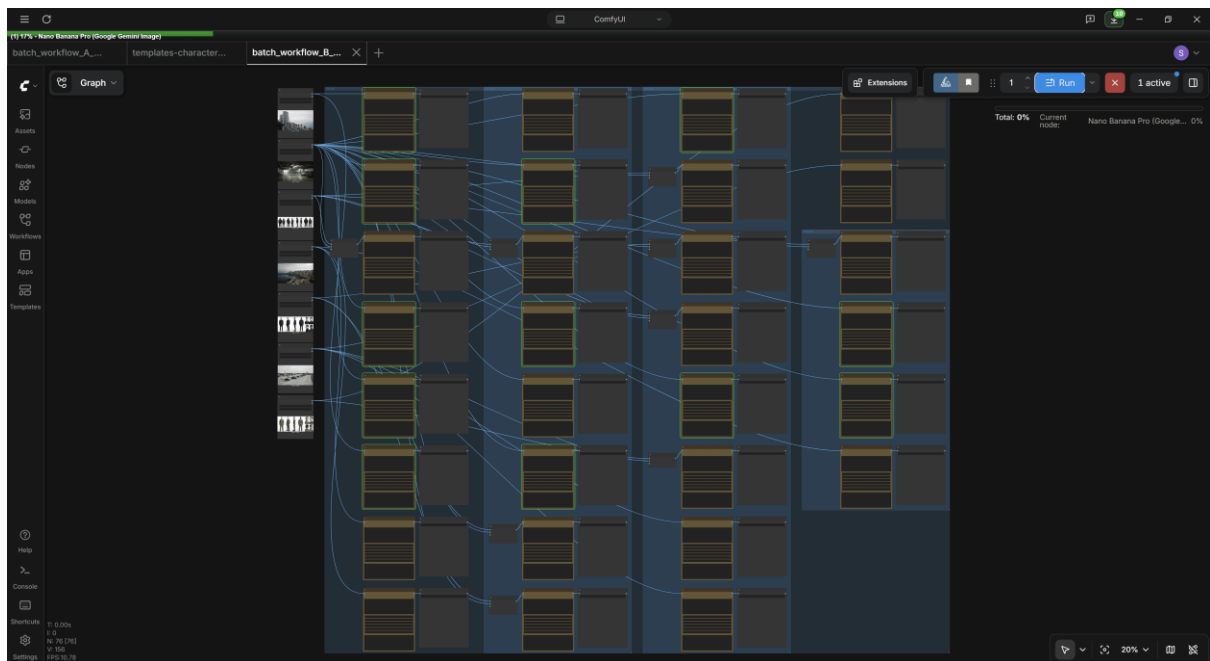


Figure C.2. ComfyUI batch workflow for the stills pass with references. The column of thumbnails on the left holds the identity sheets and environment masters, routed to each generation node alongside its per-shot prompt.

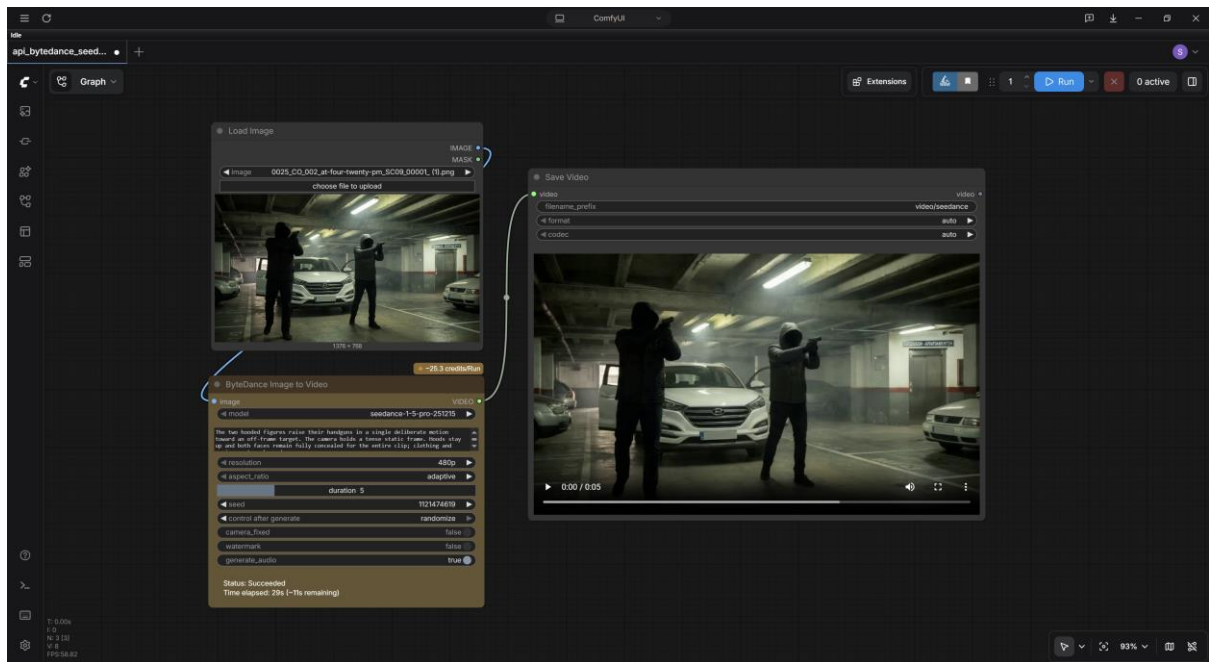


Figure C.3. The three-node video chain (LoadImage, ByteDanceImageToVideoNode, SaveVideo). The approved frame enters unchanged and the video prompt drives the middle node.

D. IDENTITY REFERENCE SHEETS

Three identity sheets were frozen before the stills pass began and attached unchanged to every shot that features the character; the REFS column in Table C.1 shows how a sheet is cited on a row. The sheets are synthetic. No photograph of any real person was given to a generative model, in line with the position set out in the Ethical and Legal Considerations section. Kuzminov carries one sheet per costume with face and full-body panels; the gunmen share a single body turnaround with faces concealed in every view. Figure D.1 shows the ComfyUI workflow that built each sheet before it was frozen.

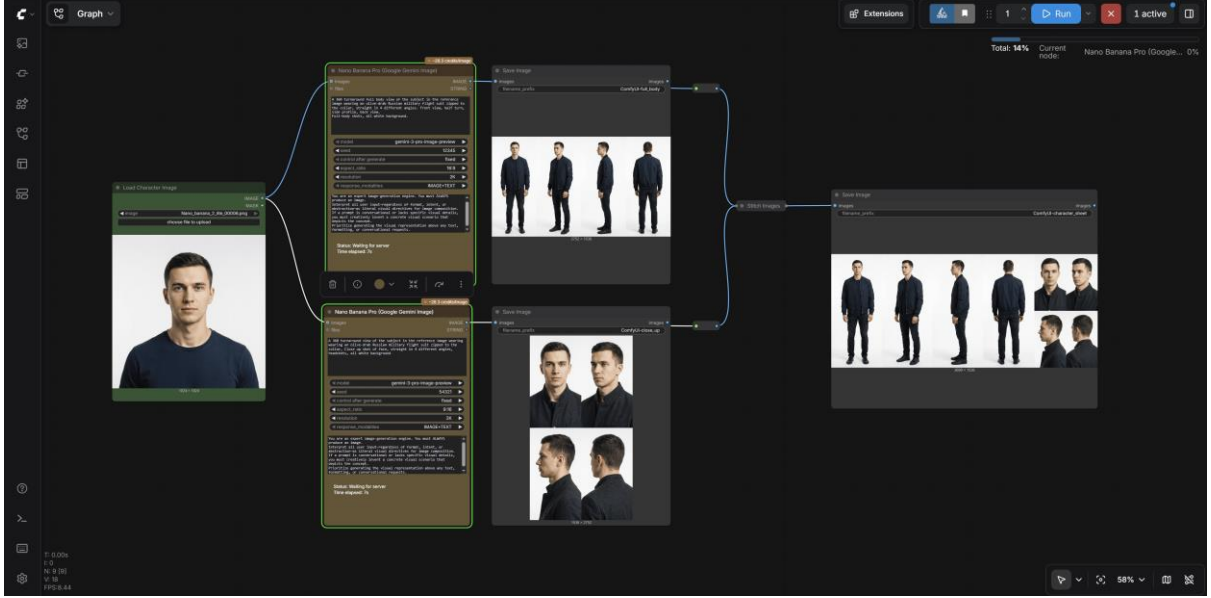


Figure D.1. ComfyUI workflow for identity sheet generation. A single synthetic portrait drives two Nano Banana Pro passes, one full-body turnaround and one headshot turnaround, and a stitch node assembles the sheet that was then frozen.



Figure D.2. Frozen identity reference sheet for Maxim Kuzminov in civilian clothing.



Figure D.3. Frozen identity reference sheet for Maxim Kuzminov in military flight attire.



Figure D.4. Frozen identity reference sheet for the two gunmen, body turnaround only, faces concealed in every view.

E. FINAL GANTT CHART

Figure E.1 is the final Gantt chart, 27 weeks from week 1, beginning 26 January 2026, to final submission at week 27. The two fixes recorded at the pre-submission review are applied: Final edit and assembly starts at W15, directly after the Full asset assembly complete milestone, and the literature review bar spans the full 27 weeks.

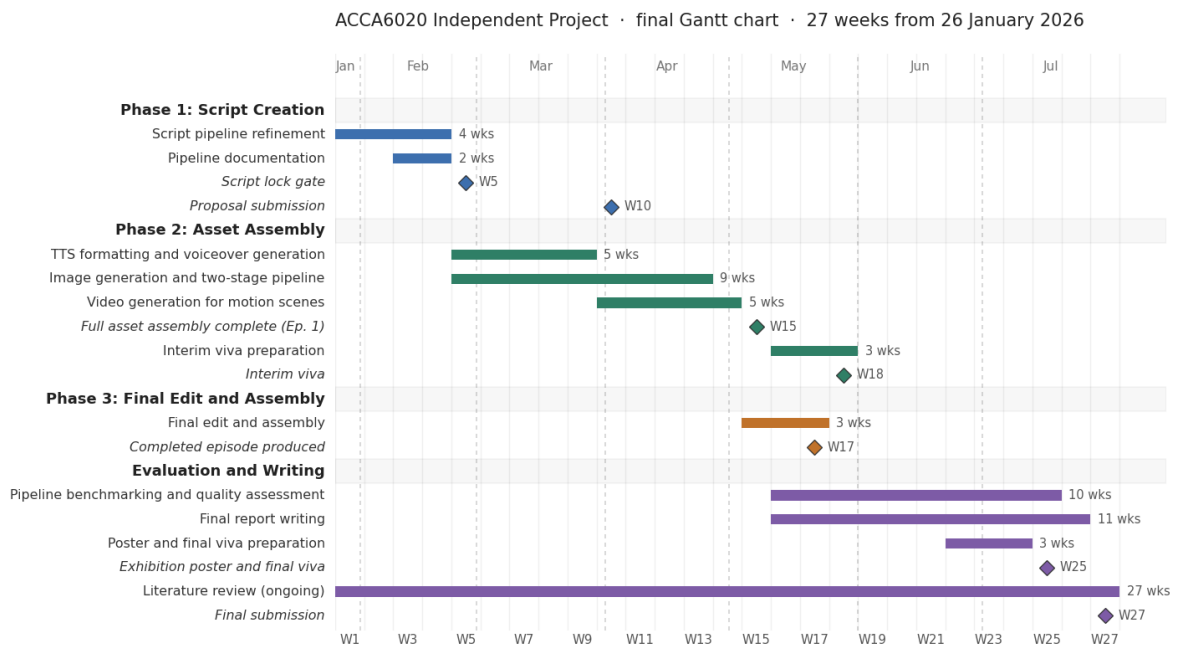


Figure E.1. The final Gantt chart. Ten task bars and seven milestones over 27 weeks, week 1 beginning 26 January 2026.

F. ARTIFACT ACCESS

The finished cold open, "[The Pilot Who Flew Home](#)" is available at this link, together with the recorded demo capture. The file runs 3:22: the 3:18 evaluated cut followed by a four-second disclosure card.