Synthetic Studios

Synthetic Studios

AWC Field Manuals No. 2

Abdul Kavungal (AK)

AWARD WINNING CREATIONS

Synthetic Studios
AWC Field Manuals No. 2

Copyright © 2026 Abdul Jaleel Kavungal. All rights reserved.

Published by Award Winning Creations.

No part of this book may be reproduced or transmitted in any form without written permission from the publisher, except for brief quotations in reviews and the templates in Appendix A, which readers may adapt for their own use.

Tool, model, price and policy details are as checked in November 2026 and are date-stamped in the text. They will change. Statements attributed to companies, unions and public bodies are reported as those organisations published them; sources are listed in the Notes and References. Nothing in this book is legal advice.

Kestrel Lane Pictures, and its films *Low Tide* and *The Salt Road*, used throughout as running examples, are composite and illustrative. They do not exist, and their budgets, schedules, shot counts and decisions are illustrative, not data about any real company or production.

Product and company names mentioned are trademarks of their respective owners. The information in this book is provided without warranty. Rights, labour and disclosure decisions should be made with appropriate professional and legal advice.

Typography by TyAK: Keyslate Display and TyCode, under the SIL Open Font License 1.1. Body text set in Source Serif Pro.

Contents

Preface vii

PART I. FIRST LIGHT **1**

1. The Studio in the Machine 3
2. Your First Film in Thirty Days 11
3. How the Models Work, Enough to Direct Them 19

PART II. PRE-PRODUCTION **29**

4. Story First: Writing in the Age of Generators 31
5. Look Books, Storyboards and Previs 39
6. Casting the Synthetic and the Human 47

PART III. PRODUCTION **57**

7. Directing the Generator 59
8. Camera, Motion and Control 67
9. Capturing the Real: Photogrammetry, NeRFs and Gaussian Splats 75
10. Virtual Production and the LED Volume 83
11. Performances, Faces and Digital Humans 93

PART IV. POST-PRODUCTION **101**

12. Editing the Synthetic Film 103
13. AI VFX: Roto, Relighting, Cleanup and Upscaling 111
14. Voice, Dialogue and Music With Consent 119
15. Colour, Finishing and Delivery 127

PART V. THE STUDIO PIPELINE **137**

16. Node Graphs and Agentic Pipelines 139
17. Cost, Schedule and Crew 149

PART VI. RIGHTS, LABOUR AND THE FRONTIER **159**

18. Rights, Copyright and Clearance 161
19. Labour, Consent and Ethics 171
20. Provenance, Disclosure and Distribution 179
21. World Models and the Road Ahead 187

Appendix A. Templates 195
Appendix B. A Ninety-Day Plan 199
Glossary 203
Notes and References 209

Index 217

Preface

This book began with a simple observation: the people asking how to make films with AI were asking two very different questions, and most of the advice available answered neither of them well.

The first question came from people who had never finished a film. They had seen a striking generated clip, tried a few prompts, and found themselves with a folder of eight-second fragments that did not add up to anything. They wanted to know how to get from a pile of shots to a finished piece of work that someone would watch to the end.

The second question came from people who had finished many films. Editors, producers, VFX supervisors and directors were watching tools arrive every month and trying to work out which ones belonged in a real pipeline, which ones would survive a legal review, how to price a job that used them, and what their unions, insurers, distributors and festival programmers would say. They wanted to know how to build a studio, not how to write a prompt.

These two questions have more in common than they seem. Both are about turning generation into production. A generated clip is raw material, in the same way that a roll of exposed negative was raw material in 1895. What makes it a film is everything around it: the story it serves, the choices about which take to keep, the continuity from shot to shot, the sound, the colour, the consent of the people whose faces and voices appear, the record of how it was made, and the delivery that gets it in front of an audience. That surrounding system is what this book calls a synthetic studio.

Who this book is for

If you are starting out, read Part I closely and do the thirty-day project in Chapter 2. It is designed so that a single person with a laptop and a modest monthly budget can finish a ninety-second short, with sound,

with a clean delivery file, and with honest disclosure of how it was made. Then work through Parts II to IV at whatever pace your projects demand.

If you already work in film or television, skim Chapter 2 and start with Chapter 3, which explains the models in enough depth to direct them, then move to whichever later chapters match your department. Parts V and VI are written for you: node-based and agentic pipelines, cost and crew, rights and clearance, labour agreements, provenance and distribution, and the frontier of world models, volumetric capture and digital humans.

What this book covers

Part I, First Light, sets out what changed, walks through a first finished film, and explains how image and video models work.

Part II covers pre-production: writing, look development, storyboards and previsualisation, and the casting decisions that determine whether a character looks like the same person from one shot to the next.

Part III covers production, whether the camera is real, virtual or generated: directing a video model shot by shot, controlling camera and motion, capturing real places as Gaussian splats and radiance fields, shooting on LED volumes driven by real-time engines, and working with performance capture, lip-sync, de-ageing and digital humans.

Part IV covers post-production: editing, AI-assisted visual effects, voice and music with consent, colour and finishing, and the technical specifications a distributor will check.

Part V is about the studio as a pipeline: node graphs, agents that operate tools on your behalf, asset and version management, cost, schedule and crew.

Part VI covers rights, labour and the frontier: what copyright protects in an AI-assisted film, the litigation and licensing that shape the tools, what the 2023 and 2026 guild agreements require, how provenance standards and disclosure rules work, and what world models and volumetric cinema may change next.

How to read the claims in this book

This is a fast-moving field, and a book about it has to be honest about time. Every statement about a specific tool, model, price or feature carries a date, usually "as of November 2026." Some of those statements will be out of date by the time you read them. Models are released, renamed, merged and withdrawn; one of the best-known video generators of 2025 was shut down in 2026, and a feature film that depended on it had to change course. The principles in this book are meant to survive those changes. The product details are there so you can see the principles applied, and so you can check them against what you find today.

No statistic, quotation or citation in this book is invented. Where the book reports a figure, it reports it as its source published it, and the Notes and References at the back list every source by chapter. Where a claim comes from a company describing its own product, the book says so. Where the book describes Kestrel Lane Pictures and its short film *The Salt Road*, it is describing a composite, illustrative studio assembled from common practice. Kestrel Lane does not exist, and nothing said about it refers to a real company or production.

Nothing in this book is legal advice. Rights, labour and disclosure rules differ by country and change often. Use the chapters in Part VI to ask better questions, then ask them of a qualified lawyer, your union or guild, and your distributor.

A note on the series

Synthetic Studios is the second of the AWC Field Manuals. The first, *Software Factories*, argued that when AI starts producing software you are running a production system, and production systems need design, boundaries, evidence and accountability. The same is true when AI starts producing images. The third manual, *Signal Rooms*, goes deeper into AI sound engineering than Chapter 14 of this book can, and the two are written to be read together.

The cover machine for this volume is the Lumière Cinématographe of 1895, which recorded, projected and printed film in one box. It is a good emblem for the subject. The Lumières did not invent cinema by building a better camera alone. They built a system that took an image all the way from capture to audience. That is still the job.

PART I

First Light

CHAPTER 1

The Studio in the Machine

On 28 December 1895, in a basement room of the Grand Café on the Boulevard des Capucines in Paris, a paying audience watched projected moving pictures made with a machine called the Cinématographe. Auguste and Louis Lumière had patented it in February of that year. It was a wooden box with a hand crank, a lens and an intermittent mechanism that pulled film past an aperture one frame at a time. What made it remarkable was not any single part. It was that the same box could record film, project it to an audience, and, with a light source behind it, print positive copies from the negative. Capture, duplication and exhibition fitted into something a person could carry.

The films were short, under a minute each, and simple: workers leaving a factory, a train arriving at a station, a baby being fed. Within a few years other filmmakers had worked out that the moving image could do much more than record. Georges Méliès used stop tricks, multiple exposures and painted sets to make people vanish and rockets land in the eye of the moon. Editors discovered that two shots placed side by side created a meaning neither had alone. The machine was the beginning. The craft was what people built around it.

That is a useful frame for what happened to filmmaking between 2024 and 2026. Machines arrived that could generate moving images from a sentence, a photograph or a rough performance recorded on a phone. Within two years they went from producing a few seconds of dreamlike, unstable footage to producing clips of fifteen or thirty seconds with synchronised dialogue and sound, consistent characters drawn from reference images, and resolutions that a colourist could work with. People made films with them, some of which played at festivals, some of which reached streaming services, and many of which nobody watched to the end.

This book is about the difference between those outcomes. It argues that the machine is not the studio. A video model is closer to the Lumières' box than to the Lumières' company: a remarkable device that captures, or in this case synthesises, the raw material. The studio is the system around it.

What changed, and when

It helps to be precise about the steps, because the change came quickly and the steps blur.

The first step was the still image. By 2023, diffusion models could produce photographic or painterly images from text descriptions, and filmmakers began using them for mood boards, concept art and storyboards. This was a pre-production change. The final film was still shot or animated the usual way.

The second step was short video. Through 2024, a series of models from research labs and start-ups could generate a few seconds of video from a text prompt or animate a still image. The results were striking in isolation and difficult to edit together. Characters changed faces between shots. Hands melted. Physics was approximate. Most of the footage that worked was atmospheric: landscapes, abstract motion, establishing shots where nobody had to act.

The third step, which ran through 2025, was control. Video models gained the ability to take reference images of a character, location or object and keep it recognisable across separate generations. Runway introduced references with its Gen-4 model in March 2025; Kling had introduced a multi-image "Elements" feature in January of that year. Models gained first-frame and last-frame control, so a shot could be made to begin and end on chosen images. Performance-capture tools arrived that transferred an actor's recorded face and body movement onto a generated or illustrated character. Video-to-video editing models such as Runway's Aleph, announced in July 2025, could take existing footage and change the lighting, add or remove objects, or generate a new angle on the same moment. Models also began to generate synchronised sound and dialogue with the picture.

The fourth step, through 2026, was consolidation and industrialisation, and it was messier than the earlier ones. Clip lengths grew: by November 2026, several leading models could generate fifteen to thirty seconds in one pass, according to their makers and to independent comparisons. Open-weight models capable of synchronised audio and video, such as Lightricks' LTX-2, released in January 2026, made it possible to run serious generation on local hardware and to fine-tune it. Node-based tools such as ComfyUI added the major commercial models as nodes, so a studio could build a repeatable graph instead of clicking through a web interface. AI agents began to operate those graphs through standard protocols. Interactive world models, which generate an explorable environment in real time rather than a fixed clip, moved from research announcement to limited public access.

At the same time, the industry discovered how fragile it was to build on a single vendor. OpenAI's Sora, one of the most discussed video models of 2024 and 2025, was shut down: the app and website closed in April 2026 and the API in September 2026. An animated feature that had been made with Sora had to move to other models and missed its planned festival debut as a result. That episode belongs at the start of this book, because it teaches the most practical lesson in it: design your studio so that any single model can be replaced.

Three kinds of AI filmmaking

People use "AI filmmaking" to describe at least three different practices, and confusion between them causes a lot of bad advice.

The first is **fully generated** filmmaking. Every frame on screen comes from a generative model. There may be no camera, no set and no actors, although there is still writing, direction, editing, sound and finishing. Most short AI films shared online are of this kind, as are many of the films shown at dedicated AI film festivals.

The second is **hybrid** filmmaking. Real performances, locations or plates are captured with cameras, and generative tools are used for some shots, some elements within shots, or some departments. The Tribeca 2025 short *Ancestra*, directed by Eliza McNitt and produced by Darren Aronofsky's Primordial Soup with Google DeepMind, combined live-

action performance with generated imagery. When Netflix said in July 2025 that it had used generative AI for a building-collapse sequence in the Argentine series *El Eternauta*, it was describing a hybrid use inside an otherwise conventional production.

The third is **AI-assisted** filmmaking, in which generation does not produce final images at all but supports the people who do: script analysis, storyboards, previsualisation, rotoscoping, relighting, upscaling, transcription, translation, temporary music and voice for an edit, scheduling and budgeting. Much of the measurable change in professional film and television has been of this kind, because it fits existing pipelines and existing agreements most easily.

Most real studios will do all three, on different projects or in different departments of the same project. The chapters that follow address each. The important point is that the legal, labour and quality questions are different in each case. A fully generated film raises sharp copyright questions because there may be little human-authored expression in individual frames. A hybrid film raises questions about consent, because real performers may be altered. An AI-assisted film raises questions about data handling and about whose work is being displaced.

A definition

This book uses the following definition:

> *A synthetic studio is a production system that turns stories into finished, deliverable films using generative models alongside conventional capture, with defined stages, controls on consent and rights, a record of how every shot was made, and people accountable for the result.*

Each part of the definition matters.

A production system means repeatable stages with inputs and outputs, not a sequence of inspired accidents. A system can be improved, measured and handed to another person.

Turns stories into finished, deliverable films puts the emphasis on finishing. A studio is judged by what it delivers. A shot that cannot be cut into a sequence, matched in colour, cleared for rights or exported to a distributor's specification is not an asset. It is a sketch.

Alongside conventional capture recognises that the most capable studios of 2026 are hybrid. Cameras, microphones, performers, sets and real places remain the most reliable source of believable motion, performance and light. Generation extends what a small crew can put on screen. It rarely replaces everything.

Controls on consent and rights names the two areas where a synthetic studio carries risks that a conventional one mostly does not. A model can produce a face that resembles a real person, a voice that resembles a real performer, or a composition that resembles a copyrighted work, without anyone intending it. A studio needs explicit checks for each.

A record of how every shot was made is provenance. It serves several purposes: it lets you regenerate or repair a shot later, it supports copyright registration by showing where human authorship lies, it answers questions from distributors and insurers, and it lets you attach content credentials to the finished film.

People accountable for the result is the part that does not change. The director is still responsible for what is on screen. The producer is still responsible for the budget, the schedule, the contracts and the clearances. Generation changes how the work is done, not who answers for it.

The argument of this book

Five claims run through the chapters that follow.

First, story and performance still decide whether a film works. A generated image can be beautiful and empty. Audiences forgive rough pictures far more readily than they forgive confusing stories or dead performances. The AI-assisted films that hold an audience tend to spend their effort on writing, pacing, sound and a small number of shots that matter, and keep everything else simple.

Second, control beats novelty. Every year a new model produces more impressive single clips. What a filmmaker needs is not the most impressive clip but the ability to get a specific shot, matching the shots around it, on purpose and repeatably. That is why so much of this book is

about references, camera control, motion transfer, capture and pipelines.

Third, capture is coming back. It is tempting to think of generation as the opposite of photography. In practice, the frontier in 2026 combined them: Gaussian splats of real locations dropped into real-time engines, actors' performances driving generated characters, live footage relit after the fact from estimated material properties, and LED volumes showing generated or captured worlds behind real performers. Real-world capture gives generation something true to hold on to.

Fourth, consent and provenance are production requirements, not compliance afterthoughts. The guild agreements of 2023 and 2026, state and national laws on digital replicas, the EU's transparency obligations that took effect in August 2026, studio guidance for production partners and platform disclosure rules all assume that a production knows which elements were generated, whose likeness and voice appear, and what permission exists. A studio that records this as it works pays almost nothing for it. A studio that tries to reconstruct it at delivery pays a great deal.

Fifth, design for replacement. Models, prices and terms change month by month. A studio's durable assets are its stories, its characters' reference sets and trained adapters where the licence allows, its captured locations, its templates, its graphs and its records. Treat any specific model as a replaceable component behind those assets.

What we know and what we do not

A book about a fast-moving field should be clear about the limits of its evidence.

We know a good deal about what the tools can do, because their makers publish descriptions and examples, independent leaderboards compare them, and many people use them in public. We know about a small number of professional productions in some detail, because studios, platforms and filmmakers have chosen to describe them: the generated building collapse in *El Eternauta*, the hybrid short *Ancestra*, the real-time de-ageing on *Here*, the visual dubbing of *Watch the Skies*. We know what

the major labour agreements, copyright guidance and disclosure rules say, because they are published.

We know much less about how widely generative tools are used in professional production overall, how much they save, and what they do to the quality of the work. Many productions use AI tools without saying so, and those that disclose tend to be the ones with good stories to tell. Claims that AI cuts production costs by some large fraction, or that it will replace some category of worker by some date, should be read with that in mind. This book avoids such figures because the evidence offered for them is usually thin.

What the book offers instead is practice: how to work, assembled from published sources, from the documented behaviour of tools, and from the long experience of film production, which has absorbed new technologies many times before. Where it describes a practice as common, it means common in the published material and tool documentation, not measured across the industry.

Kestrel Lane Pictures

To keep the chapters concrete, the book follows a composite, illustrative studio called Kestrel Lane Pictures. Kestrel Lane is two people: a director-writer who used to cut commercials, and a producer-editor who came from documentary. They have a modest equipment budget, a good microphone, a mid-range camera, a workstation with a single consumer graphics card, and subscriptions to two or three generation services at a time.

In Chapter 2 they make their first project, a ninety-second short. Over the course of the book they develop *The Salt Road*, a twelve-minute hybrid short set on a coastal highway, with two live-action performers, a generated sea-storm sequence, a captured lighthouse interior, a digitally extended landscape and a score that mixes recorded and generated material. Later chapters follow them as they take on commissioned work and build the pipeline to handle it.

Kestrel Lane does not exist. Its budgets, shot counts and decisions are illustrative, chosen to show how the practices in this book fit together, and they are not data about any real company.

What this book will not do

It will not rank models. Rankings in this field change faster than a book can be printed, and the right model for a shot depends on the shot. Where the book names models, it does so to illustrate a capability as of a stated date.

It will not teach prompt "secrets." There are no magic words. There is clear description of what the camera sees, what moves, how the light falls and what the shot is for, which is what a good director has always given a crew.

It will not pretend the hard questions are settled. Whether training generative models on copyrighted films and images without permission is lawful was still being litigated in several countries in November 2026. Whether audiences will accept fully generated performers is an open question, and the main performers' union in the United States had made its position plain. The book reports where things stand and suggests how a careful studio can work while they remain unresolved.

How the rest of the book is organised

Chapter 2 is a complete first project. Chapter 3 explains how the models work in enough depth to direct them. Part II moves through pre-production, Part III through production with generated, captured and virtual cameras, and Part IV through post-production and delivery. Part V turns the practice into a pipeline. Part VI covers rights, labour, provenance, distribution and the frontier.

At the back you will find templates you can copy, a ninety-day plan for setting up a studio, a glossary, the notes and references, and an index. The templates are deliberately plain. The fields matter more than the format.

The Lumières' box was valuable because it carried an image all the way from the world to an audience. A synthetic studio has the same job. The difference is that some of the world is now made inside the box, and you have to be able to say which parts.

CHAPTER 2

Your First Film in Thirty Days

The fastest way to understand a synthetic studio is to finish something. This chapter walks through a complete first project: a ninety-second short film, made by one or two people over thirty days of part-time work, with picture, sound, titles, a clean delivery file and an honest note about how it was made. It is deliberately modest. The aim is not to make a masterpiece. It is to go through every stage once, so that the rest of the book has something to attach to.

Kestrel Lane Pictures, the composite studio followed throughout this book, made this kind of project before attempting anything larger. Their first short was called *Low Tide*: a woman walks out across exposed sand at dawn to find something the sea has left. It has no dialogue, one character, four locations and a voice-over of three sentences. The constraints were chosen on purpose, and you should choose yours the same way.

Choose constraints that suit the tools

Every tool has strengths and weaknesses, and a first project should play to the strengths.

As of late 2026, video models were strongest at single shots of moderate length with clear subjects, readable light and simple motion: a person walking, a camera pushing in on a face, waves breaking, clouds moving, a vehicle on a road. They had improved greatly at keeping one character consistent across shots when given references, and several could produce synchronised dialogue. They remained weaker at precise hand interactions with objects, at long continuous action with several people, at legible text in the frame, at exact choreography and at keeping small details stable through many shots.

So a good first project has:

- one or two characters, not five;
- little or no on-screen dialogue, or dialogue that can be delivered as voice-over;
- locations that can be described clearly and that do not depend on exact geography;
- a story that can be told in twelve to twenty shots;
- a strong sound idea, because sound carries more of a short film than beginners expect.

Low Tide met every one of these. The single character could be established with a small reference set. The beach, the dunes, a car park and a rock pool were easy to describe. The voice-over meant no lip-sync. The story turned on a single object and an expression.

What you need

The kit for a first project is modest:

- a computer able to run a modern web browser and an editing application smoothly; a dedicated graphics card helps but is not essential if you use online services;
- subscriptions to one image generation service and one video generation service, chosen after reading their terms for commercial use if you might ever show the film publicly; many offer monthly plans that can be cancelled after the project;
- an editing application; several capable ones are free;
- a decent microphone and headphones, or access to a quiet room and a borrowed microphone;
- a phone for recording ambience and reference;
- a spreadsheet or document for the generation log.

Set a budget for generation before you start, and track it. Credits disappear quickly when a difficult shot is repeated, and a fixed budget is a useful discipline: it forces the decision to redesign a shot rather than retry it.

Set a time budget too. Thirty days of part-time work, perhaps an hour or two on weekdays and longer at weekends, is enough for a ninety-second film if the scope is right. If you find yourself at day twenty with half the shots made, cut shots rather than extend the schedule. Finishing is the point.

Days 1 to 3: the one-page script

Write the film before you generate anything. A ninety-second film is roughly one and a half pages in standard screenplay format, but for a first project write a single page of plain prose describing what the audience sees and hears, in order. Each paragraph should be roughly one shot.

Write it yourself. There are good reasons to use language models for some development work, discussed in Chapter 4, but on a first project the point is to discover what you want to say. A generated script will give you a competent average of what many people have said, and you will spend the rest of the project trying to make it feel like yours.

When the page is done, read it aloud and time it. Cut anything that does not change what the audience knows or feels. *Low Tide* lost a scene in a café at this stage because it explained something the final shot could show.

Days 4 to 6: the look and the reference set

Next, decide what the film looks like, and fix it in a small set of images.

Create a look book of ten to twenty images. You can use an image model for this, or photographs you take, or frames from films you admire for private reference. Look for decisions you can describe in words: the time of day, the colour of the light, the contrast, the lens (wide and close, or long and compressed), the texture (clean digital, grainy, soft), and what the frame leaves out.

Then build the **reference set** for your character. This is the most important single asset in the project. It is a small group of images, usually generated with an image model, showing the character from the front, in three-quarter view and in profile, in the costume they wear in

the film, under neutral and under film-like light. Most video models of 2026 accept one or more reference images and try to keep the subject recognisable. The better and more consistent your reference set, the less time you will spend fighting drift later.

Two cautions apply from the first day. Do not build a character from a photograph of a real person unless that person has given clear permission for this specific use; Chapter 6 and Chapter 19 explain why. And do not prompt for the style or appearance of a named living artist, actor or franchise. Apart from the legal and ethical problems, it makes your film look like a copy.

Kestrel Lane made their character with an image model, then edited the best result by hand to remove details that kept changing: an earring, a pattern on a jacket. Simplifying a character's design is one of the cheapest ways to improve consistency.

Days 7 to 9: shot list and storyboard

Turn the page into a numbered shot list. For each shot, write:

- what the camera sees (subject, setting, framing);
- what moves (the subject, the camera, the environment);
- how long it should last on screen;
- what the shot is for in the story;
- the sound you will hear.

Then make a rough storyboard by generating or drawing a single still for each shot. These stills can become your starting frames: many video models generate from an input image, and a carefully composed still gives you much more control than text alone. Chapter 5 covers storyboards and previsualisation in more depth.

Low Tide had sixteen shots, with an average screen duration of a little under six seconds. That is a reasonable density for a short film. It means generating perhaps sixteen good shots, which in practice means generating many more.

Days 10 to 18: generating the shots

Now generate. Work shot by shot, in story order, and keep notes.

For each shot, start from the storyboard frame and your reference images, then write a short, concrete description of the motion: "She walks slowly away from the camera across wet sand toward the waterline. The camera stays still at waist height. Low sun from the left, long shadows. Wind moves her coat." Avoid adjectives that do not describe anything visible. "Cinematic" and "beautiful" tell a model very little; "low sun from the left, long shadows" tells it a great deal.

Generate several variations of each shot. Review them at full size and at full speed, not as thumbnails. Look for:

- whether the character still looks like the character;
- whether anything impossible happens (extra limbs, objects merging, sudden changes);
- whether the motion feels like the motion you wanted;
- whether the first and last moments are usable as cut points.

Keep the best one or two. Write down which model, which settings, which input images and which prompt produced them. This record is your **generation log**, and it becomes important later for repair, for disclosure and for copyright. A template is in Appendix A.

Expect a ratio of attempts to usable shots. In conventional filmmaking, the shooting ratio is the footage shot divided by the footage used. Generation has the same idea, and the ratio varies widely with the difficulty of the shot and the model. Simple atmospheric shots may work on the first or second try. A shot with a character interacting with an object may take many attempts, or may need to be redesigned. Budget your credits and time accordingly, and when a shot keeps failing, change the shot rather than repeating the prompt. That is the most important practical lesson in this chapter, and Chapter 7 returns to it.

Kestrel Lane's hardest shot was the moment the woman picks up the object from the sand. After many failures they replaced it with two simpler shots: a close-up of her hand reaching down, ending before contact, and a wider shot of her standing up with the object already in

her hand. The cut did the work that the model could not. Editors have solved problems this way for a century.

Days 19 to 23: the edit

Bring the selected shots into an editing application. Any standard editor will do. Lay them out in script order and watch the whole thing.

It will not work yet. Some shots will be too long and some too short. The rhythm will be wrong. A shot you liked in isolation will feel out of place. This is normal, and it is the stage at which a film is actually made.

Trim each shot to the moment where it does its job. Generated shots often have a usable middle with weaker beginnings and endings, where motion is starting up or slowing down. Use the middle. If a shot is too short, consider whether a cut to a reaction or a detail would serve better than extending it. Several editing applications now include generative extension that adds a second or two to the head or tail of a clip; it is useful for a breath of extra time, not for fixing a shot that does not work.

When the cut feels right without sound, build the sound.

Days 24 to 27: sound

Sound is half of the film. A ninety-second short with good sound and adequate pictures will hold an audience; the reverse rarely does.

Build it in layers:

1. **Ambience**: the continuous background of each location. For *Low Tide*, wind, distant surf and gulls. Record your own if you can. A phone recording of a real beach is often better than a generated one.
2. **Effects**: footsteps on wet sand, the rustle of a coat, the sound of the object. Match them to picture.
3. **Voice**: Kestrel Lane's voice-over was recorded by a friend in a quiet room with a decent microphone. If you use a synthetic voice, use a model whose terms permit your use and do not clone a real person's voice without their written consent. Chapter 14 explains why.

4. **Music**: commissioned, licensed, or generated with a service whose terms clearly permit your intended distribution. Read those terms; they differ by plan and change over time.

Mix so that the voice is always clear, then check the mix on headphones, on laptop speakers and on a phone. Measure the loudness of the finished mix with a meter that reads integrated loudness; Chapter 15 gives the common targets.

Days 28 to 29: colour, titles and export

Make the shots look like one film. Generated shots from different sessions often differ in colour, contrast and grain. A simple grade that balances exposure and colour from shot to shot, and a light, consistent layer of film grain across the whole piece, will do more for cohesion than any single beautiful shot.

Add an opening title and end credits. In the credits, list the people who worked on the film, and add a short, factual line about how it was made, for example: "Images and motion created with generative AI tools from reference designs by the filmmakers. Voice performed by [name]. Music by [name or service]." Some platforms and festivals require disclosure of this kind; all audiences appreciate honesty about it.

Export a high-quality master and a delivery copy. For a first project, a master in a high-quality intermediate codec at the project's frame rate and resolution, and a delivery copy in H.264 or H.265 at a high bit rate, is enough.

Day 30: release and record

Upload the film where you intend to show it. If you publish on a platform with a disclosure setting for realistic altered or synthetic content, use it where it applies; YouTube, for example, asks creators to disclose realistic synthetic content and labels it for viewers. If your tools attached content credentials to your exports, preserve them; Chapter 20 explains how.

Then write a one-page record of the project: what tools you used and when, what worked, what failed, what you would do differently, the total

cost and the total time. Keep it with the generation log, the reference set and the project files. This is the seed of your studio's pipeline.

What the first film teaches

Kestrel Lane's record for *Low Tide* listed several lessons that most first projects share.

Writing saved more time than any tool. Every shot that had been clearly imagined on the page was quick to generate. Every shot that had been vague took many attempts.

The reference set was the best investment. Time spent simplifying and refining the character's design paid back on every shot that included her.

Changing the shot beat repeating the prompt. The worst hours of the project came from generating the same difficult shot again and again in the hope that the next attempt would work.

Sound carried the emotion. Test viewers remembered the wind and the voice more than any image.

The record mattered. When they later wanted to make a longer version, the generation log let them find exactly how each shot was made.

The rest of this book develops each of these lessons. The next chapter explains what is happening inside the models when you generate a shot, because a director who understands the medium can ask for things the medium can deliver.

CHAPTER 3

How the Models Work, Enough to Direct Them

A cinematographer does not need to design lenses, but a good one understands focal length, aperture, depth of field and how a sensor responds to light. That understanding is what lets them ask for a shot that is possible and get it on purpose. This chapter gives the equivalent understanding for generative image and video models. It is not a technical tutorial and contains no mathematics. Its aim is to explain why the models behave as they do, so that the advice in the rest of the book makes sense and so that you can reason about new tools when they appear.

Learning a distribution

A generative model is trained on a very large collection of examples, such as images with captions or video clips with descriptions. Training adjusts millions or billions of internal numbers, called parameters or weights, until the model is good at a particular task. For most image and video generators of the past few years, that task has been some form of removing noise.

The idea, introduced for images in a 2020 paper on denoising diffusion probabilistic models and developed rapidly afterwards, is simple to state. Take a training image and add a little random noise to it. Add more, and more again, until nothing is left but static. Now train a network to reverse one small step of that process: given a noisy image, and the amount of noise added, predict the noise so it can be removed. A network that can do this well, across millions of images, has in effect learned what images in its training data tend to look like. To generate a new image, you start from pure random noise and apply the network

many times, removing a little noise at each step, until a clean image emerges.

A text description steers this process. During training, each image is paired with its caption, and the network learns to remove noise in a way that is consistent with the caption. At generation time, your prompt plays the role of the caption. This is called **conditioning**: the model's output is conditioned on the text, and, as later sections explain, on other inputs as well.

Two consequences follow, and both matter to a director.

First, the model produces something plausible, not something specific. It has learned the patterns in its training data and generates a new sample that fits those patterns and your conditions. If your description is vague, the model fills the gaps with what is most typical. This is why vague prompts produce generic images.

Second, the starting noise matters. The same prompt with a different random starting point, controlled by a number called the **seed**, produces a different result. When a tool lets you fix the seed, you can make small changes to a prompt and see their effect without everything else changing at once.

Working in latent space

Generating every pixel of a high-resolution image directly is expensive, and for video it is far more so. Most modern systems therefore work in a compressed representation called a **latent space**. A separate network, an autoencoder, learns to compress images into a much smaller grid of numbers and to decompress them back into images with little visible loss. The diffusion process runs on the compressed version, and the decoder turns the result into pixels at the end. This approach, called latent diffusion, was set out in a widely cited 2022 paper and underlies most image and video generators in use today.

For video, the compression usually covers time as well as space: a short run of frames is compressed together, so the model works on a compact block representing a few frames at once. This has visible effects. Fine, fast detail, such as rain, sparks, fingers or text, can blur or shimmer

because it is hard to represent in a compressed form. Many artefacts that people describe as the model "not understanding hands" are partly compression problems.

Transformers and the move to video

The network that removes noise was originally a convolutional design called a U-Net. Since around 2023, many systems have instead used a **transformer**, the same family of architecture that powers large language models, adapted for diffusion. A 2023 paper on scalable diffusion models with transformers showed that this design improves steadily as models and training data grow, which is part of why video quality improved so quickly afterwards.

A diffusion transformer cuts the compressed latent into small patches and treats each patch as a token, much as a language model treats a word piece as a token. For video, patches extend through time as well as across the frame. The transformer's attention mechanism lets every patch take account of every other patch, which is how the model keeps a face consistent from the start of a clip to the end, or makes the motion of a wave coherent across the frame. It is also why longer and higher-resolution clips are expensive: the number of patches, and the cost of attention between them, rises steeply with duration and size.

That cost explains several practical limits that filmmakers meet:

- **Clip length.** Most models generate a fixed maximum duration in one pass, which grew from a few seconds in 2024 to fifteen or thirty seconds for several leading models by November 2026. Beyond that, clips are extended in further passes, and each extension can drift.
- **Resolution.** Many services generate at a lower resolution and offer upscaling to 1080p or 4K as a separate step. Google, for example, described drafting at 360p or 720p and upscaling selected results when it launched Gemini Omni 1.2 Flash in November 2026.
- **Price per second.** Commercial services charge by the second of output, by credits that map to seconds, or by compute, because cost rises with duration and resolution.

Audio with picture

Several leading video models of 2025 and 2026 generate sound together with the picture: ambience, effects, music and dialogue with lip movement. In the simplest terms, the model learns a joint representation of sound and image, so that a door closing on screen comes with the sound of a door closing. This is very useful for drafts and for some finished work. For anything with important dialogue, most professional productions still record or separately generate voice, then align picture to it or it to picture, because they need control over performance, timing and rights. Chapter 14 covers this.

Autoregressive generation and world models

A different family of models generates video one frame, or one short chunk, at a time, each conditioned on what came before. This **autoregressive** approach is the same principle a language model uses to write one word after another. Its great advantage is that it can respond to new input as it goes. If, between frames, you tell the model that the camera turns left, the next frames can show what is to the left.

That is the basis of interactive **world models**. Google DeepMind's Genie 3, announced in August 2025, generated navigable environments in real time at 24 frames per second and 720p, staying largely consistent for a few minutes, according to the announcement. Genie 4, announced in October 2026, added native audio and the ability to change the world through natural-language instructions while exploring it. DeepMind's own account noted that testers in creative industries were using it as a draft mode for previsualisation, and that visual details, object permanence and physics could still drift during long sessions. Chapter 21 looks at what such models may mean for film.

The weakness of autoregressive generation is that errors accumulate. A small inconsistency in one frame becomes the basis for the next. This is why long-horizon consistency is an active research problem and why, for now, world models are better at exploration and drafting than at final pictures.

Conditioning: everything you can steer with

The most important idea for a director is that text is only one of many ways to condition a model. Over 2023 to 2026, research and products added many more. Understanding them is the key to control.

Image conditioning. An input image can serve as the first frame of a video (image-to-video), as the first and last frames with the model filling the motion between them, or as several keyframes along the way. This is the most reliable way to control composition, because a frame you have designed and approved removes most of the guesswork.

Reference conditioning. Reference images show the model a subject, such as a character, a costume, a prop or a location, without fixing where it appears in the frame. The model tries to keep that subject recognisable in whatever shot it generates. Runway described up to three references per generation for its Gen-4 model in 2025; by late 2026 some services accepted many more, and some accepted reference video and audio as well, though exact limits differ by model and change often.

Structural conditioning. A 2023 paper introduced ControlNet, a way of attaching extra inputs to an image model so it follows a depth map, an edge drawing, a human pose skeleton or a segmentation map. The same idea has since been applied to video: a model can follow the depth of a rough 3D layout, the skeleton of a recorded performance or the outlines of a previs animation. This is how a crude blocking pass in a 3D tool becomes a finished-looking shot that keeps its composition.

Camera conditioning. Research systems such as MotionCtrl, presented at SIGGRAPH 2024, and CameraCtrl, presented at ICLR 2025, showed ways to make a video model follow a specified camera path. Commercial tools now offer camera controls of various kinds, from presets (“dolly in”, “orbit left”) to explicit paths. Chapter 8 covers this in depth.

Video conditioning. Video-to-video models take an existing clip as input and transform it: changing style, lighting or environment, removing or adding objects, or producing a different angle on the same action. This is the basis of many post-production uses.

Performance conditioning. A recorded performance, often from an ordinary camera, can drive the face, body and hands of a generated or illustrated character. Chapter 11 covers performance capture.

Adapting a model: fine-tuning and LoRA

Sometimes no amount of conditioning will reliably produce a particular character, style or object, because the model has never seen anything like it. Then you can adapt the model.

Full fine-tuning, which continues training all of a model's weights on your own examples, is expensive and requires the model's weights to be available. A cheaper technique called **LoRA**, for low-rank adaptation, was introduced for language models in 2021 and quickly adopted for image and video models. Instead of changing all the weights, a LoRA trains a small set of additional weights that modify the model's behaviour. The result is a small file, often a fraction of the size of the model, that can be loaded on top of the base model to add a character, a style or a concept. A 2022 technique called DreamBooth showed how to teach an image model a specific subject from a handful of photographs, and LoRA-based variants of it became common.

For a studio, LoRAs matter for three reasons. They are the strongest consistency tool available for open-weight models, and some commercial products now offer equivalent "custom character" training; Autodesk's Flow Studio, for example, introduced character LoRA training in 2026. They are portable assets that a studio can own and version, within the limits of the base model's licence. And they carry obvious risks: a LoRA trained on a real person's face or a living artist's work without permission is exactly the kind of artifact that consent rules and copyright claims are aimed at. Chapter 6 discusses how to build and govern them.

Why generations fail

With this picture in mind, the common failures make sense.

Drift. A character's face, costume or proportions change across a clip or between clips. The model is producing a plausible sample, and

plausibility does not require identity. References, LoRAs and starting frames constrain it.

Impossible physics. Objects pass through each other, liquids behave oddly, a hand merges with a cup. The model has learned what such scenes tend to look like, not the laws that govern them. Simplify the interaction, change the angle so the contact is hidden, or capture it for real.

Mushy detail. Text, fine patterns, small faces in a crowd and fast motion degrade. Compression and the cost of high resolution are part of the reason. Keep important detail large in frame, or add it in post.

Prompt neglect. The model ignores part of a long prompt. Attention is finite, and training captions were rarely as detailed as filmmakers' prompts. Put the most important information first, describe one shot at a time, and use images rather than words for anything compositional.

Averageness. The shot is competent and lifeless. The model gravitates to what is common. Specificity in the prompt, unusual but clear reference images and strong editorial choices push against it.

Extension drift. When a clip is extended beyond its native length, colour, lighting and identity can wander. Plan cuts so that you rarely need long extensions.

The settings that matter

Most tools expose a handful of settings beyond the prompt. Their names vary, but the ideas are common, and knowing them saves a great deal of trial and error.

Seed. As described above, the seed fixes the starting noise. Fix it while you refine a prompt, so that you can see the effect of each change. Change it when you want genuinely different options.

Guidance strength. Many tools let you set how strongly the output follows the prompt and other conditions, often called guidance or prompt strength. Higher values follow the conditions more literally and can look harsh or oversaturated; lower values give the model more

freedom and can drift from what you asked. Tools usually choose sensible defaults, and it is rarely worth moving far from them.

Steps. Diffusion runs a number of denoising steps. More steps can add detail, up to a point, and take longer. Many fast or "turbo" variants of models are trained to produce good results in very few steps, trading some quality for speed. Use fast variants for drafts and full variants for finals.

Reference or image strength. When an input image or reference is used, a setting often controls how closely the output must follow it. Strong settings keep composition and identity but can freeze motion; weak settings allow more motion but more drift. For each kind of shot, find the setting that works and record it in the template.

Negative descriptions. Some tools accept a description of what to avoid. They can help remove persistent problems, such as text or a recurring object, but long negative lists tend to have unpredictable effects. Use them sparingly and specifically.

Resolution and aspect ratio. Models are trained at particular resolutions and aspect ratios and usually perform best at those. Generating at an unusual aspect ratio can lead to awkward compositions or repeated elements. If your film uses a wide aspect ratio, check which ratios your model supports natively, and consider generating at a supported ratio and framing within it.

None of these settings is a substitute for a clear shot description and good input images. They are adjustments, not solutions. Record them in the generation log, because the same prompt with different settings is a different shot.

Image models still matter

Because this book is about moving pictures, it is easy to forget how much of a synthetic studio's work happens with still-image models. Key frames, reference sets, look development, storyboards, texture and matte paintings, title designs and marketing art are all still images. In most narrative workflows, a shot begins as a still and only becomes video in its final step.

Image models are also more mature and more controllable than video models. They generally offer more precise editing (changing one region while keeping the rest), stronger structural guidance, better text rendering, more adapters and more open-weight options. A practical consequence is that the more of a shot's decisions you can make in a still image, the more control you have. Composition, character, costume, light and palette can all be fixed in the key frame; the video model is then asked only to add motion and sound.

This is why so many of the workflows in later chapters begin with an image model and end with a video model. It is also why a studio's investment in image skills, such as compositing stills, editing them precisely and keeping them consistent, pays off in its video work.

A director's mental model

Pulling this together, a useful way to think of a video model is as a very fast, very literal, extremely well-read crew that has never been to your set. It has seen a vast number of images and clips. It will produce what is typical unless you specify otherwise. It understands pictures better than words, and it follows a designed frame better than a description of one. It has no memory of your previous shot unless you show it. It cannot be held responsible for anything.

Everything that follows in this book is a way of giving that crew the information it needs, checking its work, and recording what it did, so that the people who are responsible can stand behind the result.

PART II

Pre-Production

CHAPTER 4

Story First: Writing in the Age of Generators

Every filmmaker who has worked with generative tools for any length of time arrives at the same discovery: the bottleneck is not the pictures. Once you can produce a beautiful shot of almost anything in a few minutes, the scarce resource becomes knowing which shot to produce, why, and what comes after it. That knowledge is story. This chapter covers development and writing for synthetic and hybrid productions: how to choose stories that suit the medium, how to write scripts that can be produced, where language models help and where they harm, what the writers' guild agreements say, and how to break a script down into the units a synthetic studio actually builds.

Why story matters more, not less

When photographed images were expensive, every frame had to earn its place, and that discipline pushed filmmakers toward clarity. When images become cheap, the discipline disappears unless you supply it yourself. The online stream of generated video in 2025 and 2026 showed what happens without it: long sequences of spectacular, unconnected images that audiences scroll past. A dragon over a city is no longer remarkable. A reason to care whether the dragon lands is.

There is also a structural reason. Generated shots are made one at a time, often by different models, in different sessions, with different random starting points. Nothing in the generation process connects them. The connection has to come from the script, the shot list and the edit. A film made from generated material is held together by its story to a greater degree than a film shot continuously with real actors, where the shared space and time of the set supply some continuity for free.

Choosing stories that suit the medium

Not every story suits a synthetic production, and choosing well at the start saves months.

Stories that suit generation tend to have a strong visual or conceptual premise that would be expensive or impossible to photograph: other worlds, other eras, impossible scale, dreams and memory, nature at extremes, the inside of the body, abstraction. They often rely on voice-over, narration or music rather than dense dialogue. They often have few principal characters. Many successful AI shorts have been essays, fables, memories and music-driven pieces for this reason.

Stories that suit hybrid production have human performances at their centre and use generation to extend the world around them: a real actor in a generated landscape, a real location extended with generated architecture, a real chase finished with a generated crash. These are the stories where small crews gain the most, because the performance supplies the emotional truth and generation supplies production value.

Stories that resist generation depend on subtle, sustained performances between several people in the same space, on precise physical comedy, on dense dialogue, or on fidelity to real events and real people. These stories can still use AI-assisted tools throughout post-production, but generating their key images is likely to cost more effort than shooting them, and raises consent and accuracy questions.

Kestrel Lane chose *The Salt Road* because it fitted the hybrid profile. Its two characters, a retired lighthouse keeper and a young surveyor sent to decommission the light, would be played by real actors in a small number of real interiors. The storm, the sea, the cliffs and the long coastal road, which would have cost far more than their budget to shoot, would be generated, extended or captured.

Development: what goes on the wall

Development is the stage between an idea and a script. In a synthetic studio, it produces four things.

A logline. One sentence: who wants what, what stands in the way, and why it matters. If you cannot write it, you are not ready to write a script.

A treatment. A few pages of prose telling the story in the present tense, as the audience would experience it. The treatment is where structure gets solved: where the turning points fall, what the ending means.

A world document. For synthetic productions, this is more important than in conventional ones. It describes the look, rules and texture of the film's world in concrete terms: what era, what technology, what weather, what materials, what palette, what the light is like at different times of day, what is never seen. The world document becomes the source for look-development prompts, reference images and continuity checks.

A character document. For each principal character: age, build, face, hair, clothing, posture, how they move, how they speak. This becomes the source for the reference sets and, where appropriate, trained character models described in Chapter 6.

Writing the script

Screenplays for synthetic productions are written in the same industry format as any other. A few habits make them easier to produce.

Write what the camera sees. Screenwriting has always discouraged writing what a character thinks, because a camera cannot photograph thought. For generation, the rule is stricter: a model cannot generate intention either. "She realises he has lied" must become something visible or audible: a look, a pause, an object, a line.

Keep scenes visually distinct. Locations and times of day that look different from each other help the audience follow the story and make continuity easier, because a cut between distinct spaces forgives small mismatches that a cut within one space would expose.

Mark the method in the margin, not the text. Some writers note beside each scene whether it will be shot, generated, captured or composited. Keep this out of the screenplay itself, which should read as a story, but keep it in a production draft or breakdown. It helps the producer cost the script early.

Respect the limits, but do not write timidly. Writing only what today's models can easily produce makes for a dull film and dates quickly. Write what the story needs, then solve production problems in the breakdown.

Many apparently impossible shots become easy when broken into two or three simpler ones, as Chapter 7 shows.

Where language models help

Large language models can help in development and writing in several ways that leave authorship with the writer.

Research. Asking a model to summarise the history of lighthouse automation, the working day of a land surveyor or the sound of a particular kind of storm can speed up research, provided every fact that ends up on screen is checked against a reliable source. Models state errors fluently.

Breakdowns and logistics. Extracting characters, locations, props, time of day and scene length from a finished script is tedious and well suited to a model, with a person checking the result. The same applies to drafting shot lists from approved scenes, converting between formats and preparing schedules.

Analysis. Asking a model where a draft drags, where a character's motivation is unclear, or which scenes repeat information can be a useful first reader, as long as you treat the answer as one opinion and not a verdict.

Variations for testing. Generating several alternative versions of a line or a beat, to provoke your own better version, is a common practice. The risk is that the alternatives pull your voice toward the average.

Where language models harm

Generic structure and voice. A model asked to write a scene produces a fluent, average scene. Average is the enemy of distinctive work.

Unreliable facts. Generated research can contain invented sources, misattributed quotations and plausible errors. Anything that appears on screen as fact must be verified independently.

Authorship and rights. In the United States, the Copyright Office's January 2025 report on copyrightability concluded that purely AI-generated material is not protected and that prompts alone do not

provide sufficient human control. A script that is substantially generated may therefore be weakly protected, which matters when you try to sell or finance it. Chapter 18 covers this in detail.

Guild obligations. If you write under a guild agreement, there are rules.

What the writers' agreements say

In the United States, the Writers Guild of America's 2023 Minimum Basic Agreement with the studios' bargaining organisation set out terms on artificial intelligence that have been widely discussed and imitated. According to the WGA's own summaries, under those terms AI is not a writer, and material produced by AI is not "literary material" under the agreement. A company cannot treat generated material as source material in a way that undermines a writer's credit or separated rights; if a writer is asked to rewrite AI-generated material, they are treated as the first writer. A writer may choose to use AI tools with the company's consent and in line with company policy, but a company cannot require a writer to use them. Companies must disclose to a writer if material given to them was generated or incorporates generated material.

The WGA's 2026 agreement, ratified in April 2026 with 90.38 per cent in favour according to the guild, kept those protections and added a notice requirement: if a company licenses writers' material to an unaffiliated third party for training a public, commercially available generative AI system, it must notify the guild, which can then request a discussion that may include remuneration for writers.

Writers outside the United States should check their own guild or union. Even where no agreement applies, the WGA's terms are a useful model of fair practice: disclose generated material, do not use it to dilute a writer's credit, and do not force tools on writers.

Breaking down a synthetic script

A conventional script breakdown lists, scene by scene, the cast, locations, props, wardrobe, effects and other elements needed to shoot it. A synthetic studio adds a further layer: for each shot, how it will be made.

Kestrel Lane uses five production methods, which this book uses as standard terms:

1. **Shot**: photographed with a camera, with real performers or real places.
2. **Generated**: produced by a generative model from text, images, references or other conditioning, with no photographed plate.
3. **Captured**: built from a three-dimensional capture of a real place or object, such as a Gaussian splat or photogrammetry scan, and rendered with a virtual camera.
4. **Composited**: combining two or more of the above, for example a shot performer against a generated background.
5. **Transformed**: photographed footage altered by a model, for example relit, restyled, extended, de-aged or lip-synced.

Each method has different costs, risks and requirements. Shot elements need crew, locations, permits and performer contracts. Generated elements need credits, references and review time, and have weaker copyright protection on their own. Captured elements need capture sessions and permission to capture the location. Composited elements need VFX time. Transformed elements of real performers need consent of a specific kind, discussed in Chapters 11 and 19.

The breakdown for *The Salt Road* listed sixty-eight shots. Twenty-nine were shot, mostly interiors with the two actors; nineteen were generated, mostly the storm and the sea; six were captured, all in the lighthouse lamp room, which the studio scanned with the owner's permission; eleven were composited; and three were transformed, including one in which the older actor's performance was relit to match a generated lightning flash. These numbers are illustrative, but the proportions are typical of the hybrid projects this book describes: performance shot, world generated or captured, and the two joined in post.

The production draft

From the breakdown, Kestrel Lane produces a **production draft**: the shooting script with numbered scenes and shots, each shot carrying a

method code, a reference to its storyboard frame, the reference sets it depends on, and its sound notes. This is the document that the rest of the studio works from. When a generated shot is redone, the production draft records why. When a shot changes method, for example a generated shot that keeps failing is moved to a composite, the change is recorded with a reason.

A template for a shot card, which is one line of the production draft expanded into a working record, is in Appendix A.

Developing The Salt Road

Kestrel Lane's development of *The Salt Road* shows how these steps fit together in practice.

The idea began with a photograph the director had taken years earlier of a decommissioned lighthouse. The first logline read: "A lighthouse keeper refuses to leave." It was not a story; it was a situation. The second added want and obstacle: "When a young surveyor arrives to switch off the last manned light on the coast, its ageing keeper must decide whether to help her or keep it burning through one more storm." That gave the film a question the audience could hold.

The treatment went through four versions. The first was all atmosphere, with the storm as its climax. The producer pointed out that the storm resolved nothing between the two characters. In the final version, the storm forces them to work together to keep the light running for a ship in trouble, and the keeper realises, as the ship passes safely, that the light no longer needs him. The surveyor switches it off at dawn with his permission. The storm became a means, not an end.

The world document fixed the era (the near present, a few years after automation), the coast (unnamed, cold-temperate, grey-green), the light inside the lighthouse (brass, paraffin-yellow, then cold electric), and what would never be seen (other people, except the distant ship). The character documents described the keeper's hands in detail, because the director wanted his hands in close-up throughout, and noted that the surveyor's red waterproof jacket was the film's only saturated colour.

The script was written by the director without generative tools. A language model was used afterwards for two tasks: a first-pass breakdown, which the producer corrected, and research questions about lighthouse lamp mechanisms, each answer checked against published museum material before anything appeared on screen.

Keeping the human centre

The final point of this chapter is not a technique. It is a stance.

The people who are most effective with generative tools treat them as instruments of a vision they already have, not as a source of vision. They decide what the film is about before they generate anything. They use generation to explore how to show it, not what to say. And they accept that the hardest parts of filmmaking, which are knowing what matters, finding the truth in a scene and cutting what does not serve it, remain theirs.

That stance is not sentimental. It is practical, legal and commercial. Practical, because story is what holds generated material together. Legal, because human authorship is what copyright protects. Commercial, because audiences, distributors and festivals are already saturated with generated spectacle and are looking for work that has something to say.

The next chapter turns from words to images: how to develop the look of a film, storyboard it and previsualise it before committing to production.

CHAPTER 5

Look Books, Storyboards and Previs

Before a conventional production spends money on a set, a crew and a shooting day, it decides what the film will look like and roughly how each scene will be covered. Look development, storyboards and previsualisation are the tools for those decisions. Generative AI changed all three earlier and more thoroughly than any other part of filmmaking, because still images were the first thing the models did well. By 2026 it was also changing previsualisation through real-time world models and video generation. This chapter covers how to use these tools to make decisions, which is their purpose, rather than to make pretty pictures, which is their temptation.

Look development

The look of a film is the set of consistent visual decisions that make every frame feel as though it belongs to the same world: palette, contrast, light quality, lens character, texture, framing habits, production design and costume. In a synthetic production, the look has a second job. It must be describable precisely enough that every shot, made by different models at different times, lands inside it.

Start with words. Before generating images, write a one-page look document. Describe the palette in named colours and their relationships ("desaturated sea greens and greys; warm sodium-orange only from artificial lights; skin tones kept natural"). Describe the light ("overcast and directional by day; hard practical light at night; no glamour lighting"). Describe the lens ("mostly 35mm and 50mm equivalents, shallow focus for interiors, deep focus for landscapes"). Describe texture ("fine film grain, gentle halation on highlights, no sharpening").

Describe what never appears ("no bright blue skies; no lens flares; no neon").

Then generate to test the words. Use an image model to generate frames from the look document. Where the images miss, the document is usually vague. Revise the words until they reliably produce the look, and keep both the words and the best images. The words become a reusable prompt block; the images become style references.

Mix sources. The strongest look books combine generated frames with photographs you take, colour swatches, material samples, and private reference stills from films and paintings. Do not feed other people's copyrighted images into a model as style references for final work without permission; use them only to discuss and decide.

Decide early on texture and format. Generated images tend toward a clean, slightly glossy digital look by default. If you want grain, softness, halation or a particular aspect ratio, decide now. Grain and halation are usually better added consistently in finishing than requested in each generation, because requested texture varies from shot to shot.

For *The Salt Road*, Kestrel Lane's look document ran to a page and a half. Its most useful line was a negative one: "The sea is never blue." Generated seascapes defaulted to blue and turquoise; adding that line to the reusable prompt block, together with three grey-green reference frames, fixed it.

Production design references

Look development extends to places, objects and costumes. For each location in the breakdown, create a small set of reference images: an establishing view, a few angles, details of key surfaces and props. For each principal costume, create front, side and back views on a neutral background. For key props, do the same.

These sets do three jobs. They let the director and the producer agree what the world looks like. They serve as references when generating shots. And for hybrid productions, they brief the real production designer, costume designer and props team, who build the physical versions for the shot elements. When the generated lighthouse exterior

and the real lighthouse interior need to feel like one building, the shared references are what connect them.

Keep every reference set in a named, versioned folder. When a reference changes, increment the version and record why. A surprising amount of inconsistency in synthetic films comes from different shots having been generated against different, undocumented versions of the same reference.

Storyboards

A storyboard is a sequence of drawings or images showing each shot's framing, composition and key action, in order. It is the cheapest way to edit a film before it exists.

Generative tools make storyboards fast to produce, and that creates a trap. A board that looks like a finished frame invites people to judge it as a finished frame: to discuss its lighting and texture rather than whether the shot is needed, whether it is the right size, and whether it cuts with the next one. For story work, rough boards are often better.

A practical approach has two passes.

The story pass. Make rough boards quickly: sketches, simple generated frames in a deliberately plain style, or even stick figures. Lay them out in sequence and cut them together as an **animatic**, a timed slideshow with temporary sound. Watch it. Fix the story, the coverage and the rhythm.

The frame pass. For shots that survive the story pass, make a carefully composed key frame: the exact framing, lighting and look you want. For generated shots, this key frame is often the input image for image-to-video generation, so it is worth real effort. For shot elements, it is a reference for the director of photography.

Image models have become good at following compositional instructions, especially when given a rough layout. Many tools accept a sketch or a simple 3D blocking render as structural guidance, following the depth, edges or pose of the input while applying the look. This lets you compose a frame precisely in a simple tool and then render it in the film's style.

Shot lists and coverage

A shot list turns the storyboard into a production plan. For a synthetic studio, each line includes the method code from Chapter 4, the reference sets, the key frame and the intended duration on screen.

Think about **coverage**: the set of shots that will be available to the editor for each scene. In conventional production, a scene is usually covered from several angles, giving the editor choices. In generated production, extra coverage costs credits and review time, so there is a temptation to generate only the shots on the board. Resist it for important scenes. Generate at least one alternative angle and one cutaway for each scene, so the editor has somewhere to go when a shot does not work in sequence. Cutaways to hands, objects, landscape and reaction are cheap, low-risk and invaluable.

Previsualisation

Previsualisation, or previs, is a rough moving version of a sequence, usually made with simple 3D animation, used to plan camera moves, timing, stunts and effects before production. Large productions have used it for decades for complex sequences. Generative tools have brought it within reach of small productions and have added new forms.

Animatic previs. The simplest form is the animatic described above, with timed stills. For many short films this is enough.

Generated video previs. Generating rough video versions of each shot, at low resolution and in a deliberately simple style, lets you see motion and timing before spending on final generations or a shoot. Fast, low-cost video models are well suited to this. Several services in 2026 offered draft modes at lower resolution and lower price for exactly this purpose.

3D blocking previs. For sequences that depend on precise camera moves, spatial relationships or stunts, simple 3D blocking in a tool such as Blender or Unreal Engine is still the most reliable. Characters can be mannequins; environments can be grey boxes or, increasingly, Gaussian splats of real locations, discussed in Chapter 9. The resulting previs can later drive generation directly, through depth, pose and camera conditioning, so the work is not thrown away.

World-model previs. The newest form is to explore a scene interactively in a real-time world model. When Google DeepMind announced Genie 4 in October 2026, it reported that testers in creative industries were using it as a rapid draft mode for previsualisation: a filmmaker could step into a simulated scene and run several takes in real time, changing camera angle, lighting, dialogue and action. DeepMind also stated the model's limitations clearly: detail and object permanence could drift over longer sessions or after several complex edits, and physics, spatial accuracy and long-term consistency were still active research problems. As of November 2026, access was limited to a research prototype for certain subscribers. Treat this kind of previs as a sketchbook for ideas about blocking and camera, not as a precise plan.

For *The Salt Road*, Kestrel Lane used all four forms. The whole film existed first as an animatic with temporary sound. The storm sequence, which was entirely generated, was previsualised with low-resolution generated video to test timing. The lamp-room scenes were blocked in 3D using a splat of the real room, so the director could plan camera positions before the one-day shoot. They experimented with a world model to explore ideas for the coastal road, and kept two ideas from it as storyboard frames.

Working with world-model previs

Because world models are new, it is worth describing how to use them in pre-production without being misled by them.

Prompt the space, not the story. Describe the environment, the viewpoint, the light and the sensory details, then explore. DeepMind's own prompting guidance for Genie 4 emphasised describing the scene and viewpoint and introducing one clear change at a time. Long narrative prompts tend to produce worlds that drift.

Record everything. Exploration sessions are ephemeral. Record the screen and export whatever the tool allows, so ideas can be reviewed and shared. Note the prompts.

Extract decisions, not footage. The useful outputs of a session are decisions: this camera height, this approach to the building, this time of day, this blocking. Write them down and turn them into storyboard

frames or 3D blocking. Do not expect to cut world-model footage into the film.

Check scale and geography. World models do not know the real dimensions of your location or the size of your set. A move that feels right in a generated world may be impossible on a real stage or in a captured space. Confirm every idea in previs built on real measurements or captures before committing to it.

Mind the terms. Research prototypes often have restrictive terms on what may be done with their outputs. Read them before using material in pitches or public work.

Making decisions

All of this produces a great deal of material. The discipline of pre-production is to turn material into decisions and to record them.

At the end of pre-production, Kestrel Lane holds a **lock meeting** for each sequence. The director, the producer and anyone responsible for a department watch the animatic or previs and agree on four things for every shot: its method, its key frame, its duration, and its references. Those decisions go into the production draft. Changes after the lock are allowed, but they are recorded with a reason. This sounds bureaucratic for a two-person studio. In practice it takes an hour per sequence and saves days, because it stops the endless reconsideration that cheap images invite.

Bridging generated and photographed looks

Hybrid productions face a particular look problem: the shot elements come from a real camera, with its sensor, lenses, real light and real atmosphere, while the generated elements come from a model that has learned an average of many cameras. Left alone, the two will not match. The generated shots will usually be cleaner, more saturated and more evenly lit than the photographed ones, and the difference reads instantly in a cut.

Three habits close the gap.

Shoot first where you can. If some shots will be photographed, shoot a few test frames on the actual camera and lenses, in light similar to the scenes, before generating anything final. Use those frames as style references for the generated shots. The model will then aim at your camera rather than at an average one.

Describe the camera, not the mood. Prompts that describe lens length, depth of field, the direction and quality of light and the time of day give a model something it can match. Prompts that describe mood produce the model's idea of that mood, which is usually more dramatic than your photography.

Agree a shared finishing recipe. Decide with the colourist, before production, how all material will be brought into one colour pipeline: what working colour space, what grain, what halation, what contrast curve. Chapter 15 covers this. When both kinds of material pass through the same finishing recipe, they converge.

For *The Salt Road*, Kestrel Lane shot a morning of camera tests in the real lighthouse with the real lenses before generating any of the storm material. Those test frames, graded to the agreed look, became the style references for every generated sea shot. When they later cut a generated wave against a photographed close-up of the keeper at the window, the two belonged together.

Common pre-production mistakes

Several mistakes recur in synthetic productions, and most of them begin in pre-production.

Polishing boards instead of fixing story. Generated storyboards are so attractive that teams spend days refining frames for scenes that should have been cut. Do the story pass rough and fast.

Too many looks. Exploring dozens of styles is fun and costs little, so teams keep exploring. A film needs one look. Set a date by which it is locked.

Undocumented references. A reference image that exists only in someone's downloads folder is not a reference. Name it, version it and store it where the whole team can find it.

No method decision. Shots that reach production without a decided method tend to be generated by default, because generation is the easiest thing to start. Many of them would have been cheaper and better as shot, captured or composited elements.

Forgetting sound. Animatics without temporary sound mislead. Even rough sound changes how long a shot needs to be.

Treating previs as final. Previs, especially generated or world-model previs, is a plan. It will contain errors of scale, physics and continuity. Its job is to help you decide, not to be cut into the film.

What to keep

Pre-production produces several assets that a studio should keep and version:

- the look document and its reusable prompt blocks;
- style reference images;
- location, costume, prop and character reference sets;
- key frames for each shot;
- the animatic and previs files;
- the production draft with method codes.

These are among the most valuable things a synthetic studio owns. They are largely human-authored, so they carry stronger rights than raw generations. They are model-independent: if a model disappears, the references and key frames can drive its replacement. And they are reusable across sequels, series and commercial spin-offs.

The next chapter looks more closely at the hardest of these assets to get right: the characters.

CHAPTER 6

Casting the Synthetic and the Human

Audiences forgive a great deal, but they do not forgive a character who changes face between shots. It breaks the most basic contract of narrative film: that the person we are watching is the same person. For the first years of video generation, character consistency was the problem that kept generated films looking like music videos and mood pieces rather than stories. By 2026 it had improved enough that consistent characters across a short film were achievable, but only with deliberate work. This chapter covers that work: designing characters that can be kept consistent, building reference sets and character bibles, the methods for holding identity across shots, from simple references to trained adapters, and the consent questions that arise as soon as a character looks like, sounds like or is based on a real person.

Two kinds of casting

A synthetic studio casts in two senses.

It casts **human performers**: actors whose bodies, faces and voices appear on screen in shot elements, whose performances may drive generated characters through performance capture, or whose voices are recorded for synthetic characters. Human casting brings contracts, consent, union agreements where they apply, and the irreplaceable qualities of a real performance.

It also designs **synthetic characters**: people or creatures whose appearance is generated. Synthetic characters bring a different set of obligations: making sure they do not resemble a real person who has not consented, keeping them consistent, and being honest with audiences about what they are.

Many characters are both. A synthetic face driven by a human performance, or a real actor transformed into an older or younger version of themselves, combines the obligations of each. The rules in Chapter 19 apply to all of these cases, and you should read that chapter before casting anyone or anything for a commercial project.

Designing for consistency

Some character designs are easier to keep consistent than others, and design choices made in the first week determine how much work consistency will take for the rest of the production.

Distinctive but simple. A character with a clear silhouette, a strong colour in their costume and one or two distinctive features is easier for a model to hold than one defined by subtle detail. A red coat is easier to keep than a pattern on a scarf.

Few small details. Jewellery, tattoos, logos, patterned fabrics, badges and complex hairstyles drift. Each one is a continuity risk. Keep the ones that matter to the story and remove the rest.

Stable costume. A character who wears the same clothes throughout a sequence is far easier to keep consistent than one who changes. If a costume change matters, treat each costume as its own reference set.

Age and build in the range the model knows. Models are trained on large collections of images and tend to be most reliable on the kinds of faces and bodies they saw most often. Characters outside that range, such as very old people, people with visible disabilities or people from groups under-represented in training data, can be harder to keep consistent and can be rendered with stereotyped features. This is not a reason to avoid them. It is a reason to invest more in their references, and to have the result reviewed by people who know what a respectful and accurate depiction looks like.

The reference set

The reference set introduced in Chapter 2 is the foundation of character consistency. For a principal character in a serious production, it usually contains:

- a neutral front view, three-quarter views from each side, and a profile, at head-and-shoulders size;
- full-length front, side and back views in each costume;
- a small set of expressions: neutral, smiling, distressed, and any expression central to the story;
- the character under two or three lighting conditions used in the film;
- close-ups of anything distinctive: hands, eyes, a scar, a hairstyle detail.

All views should be on plain backgrounds and in consistent, neutral light, except the lighting studies. Generate many candidates, select carefully, and correct by hand so that the views agree with each other. An inconsistency inside the reference set will be multiplied in every shot.

For a human actor who will be transformed or extended, the reference set is made from photographs of the actor, taken with their written consent for this specific use. For a synthetic character, it is generated, then checked against the risks described later in this chapter.

The character bible

The reference set sits inside a **character bible**: a single document per character that gathers everything needed to depict them consistently. A template is in Appendix A. It records:

- the character's name, role and description in words;
- the reference set, with version number;
- the reusable prompt block that describes the character;
- any trained adapters (see below), with their base model, training data and version;
- the voice: the actor, or the synthetic voice and its licence;

- the consent record, if the character is based on or performed by a real person;
- continuity notes: injuries, costume changes, props, how the character changes over the story.

The bible is version-controlled. When anything changes, the version number increases, and every shot generated after the change records which version it used.

Methods for holding identity

There are four broad methods for keeping a character consistent across generated shots, in rising order of effort and strength. Most productions combine them.

1. Description and starting frames

The simplest method uses a precise written description in every prompt and, for each shot, a starting frame that already contains the character, made with an image model using the reference set. The video model then animates a frame in which the character is already correct. This works well for short shots with modest motion. It weakens as the character turns, moves far from the camera or is partly hidden, because the model must invent views it was not shown.

2. Reference conditioning

Most leading video models of 2025 and 2026 accept reference images of a subject and try to keep that subject recognisable in a new shot, without fixing its position. Runway described this for Gen-4 in March 2025, with up to three references; Kling introduced a multi-image Elements feature in January 2025; and later models from several companies accepted larger numbers of image, video and audio references. By November 2026 some services advertised character voices that stayed consistent along with appearance. Exact limits and behaviours vary by model and change often, so check the current documentation.

Reference conditioning is the workhorse method for most productions. To use it well, choose references that show the views the shot needs. A shot of a character from behind will hold better if a back view is among

the references. Use the same reference images for every shot in a sequence. And keep references at the same version for the whole of a sequence.

3. Trained adapters

When reference conditioning is not enough, the strongest method is to train a small adapter, usually a LoRA, as described in Chapter 3, on the character's reference set. The adapter teaches the model the character's identity, so the model can generate them from angles, distances and in situations not directly shown in the references.

Training an adapter requires a base model whose licence permits it, and either open weights you can run, or a service that offers custom character training. Open-weight video models released in 2025 and 2026, including Lightricks' LTX-2, came with training tools for adapters. Some commercial products offer equivalent features: Autodesk described character LoRA training in Flow Studio in September 2026, trained once per character and then applied to later generations.

Practical guidance for adapter training:

- **Training data.** Use your curated reference set, extended with additional consistent views. Tool guides commonly suggest starting with a few dozen good, varied, consistent images; check the current guidance for your model and trainer. Quality and consistency matter more than quantity.
- **Captioning.** Describe each training image accurately, including what is not part of the identity, such as background and lighting, so the adapter learns the character and not the setting.
- **Testing.** Generate a standard test grid after training: the character in several framings, lights, costumes and actions. Compare against the bible. Keep the test grid with the adapter.
- **Versioning.** Store the adapter, its training data, its settings, the base model version and the test grid together. An adapter is tied to its base model; when the base model changes, the adapter usually needs retraining.

4. Performance and 3D anchoring

The final method anchors identity in something outside the model: a real performer, or a 3D model of the character. A recorded performance driving a character, as discussed in Chapter 11, supplies consistent motion and expression. A 3D model of the character, rendered as a rough pass and then restyled by a video model with structural conditioning, supplies consistent shape from any angle. These approaches demand more skill and setup, but they give the most control, and they are how most high-end productions approach digital characters.

Training a character adapter: a worked example

For *The Salt Road,* Kestrel Lane trained one adapter, for a character who does not appear in person: the keeper's late wife, seen only in three photographs on the lamp-room wall and in one brief memory shot. The photographs had to look like the same woman at different ages, which reference conditioning alone did not achieve reliably.

The character was designed by an illustrator the studio hired, who drew a set of portraits at three ages. These drawings were the human-authored foundation of the character, and the illustrator was credited and paid for them, including for their use as training data. An image model, guided by the drawings, produced a larger set of photographic views at each age, which the illustrator and director reviewed and corrected until the set was consistent.

The training set had several dozen images, captioned to describe age, expression, clothing and light separately from identity. The adapter was trained on an open-weight image model whose licence permitted commercial use, on the studio's own workstation. After training, a standard test grid of the character at each age, in four framings and three lights, was generated and compared against the drawings. The first adapter overfitted to one hairstyle; a second, trained with more varied hair in the data, passed.

The adapter, its training data, captions, settings, base model version and test grid were stored together as one asset. The memory shot was then made as a video from an adapter-generated key frame. Because the

adapter was tied to the image model and not the video model, the video model received the key frame and references rather than the adapter itself, a common arrangement when the strongest video model and the most adaptable image model are different.

Continuity review

Whichever methods you use, check every generated shot against the character bible before it enters the edit. A simple review checklist includes face shape, eye colour, hair, skin tone, age, build, costume, props, distinctive marks and handedness. Review at full resolution and full speed. Drift often appears only in motion or in the last second of a clip.

Kestrel Lane's editor keeps a **continuity strip** for each sequence: one frame from each shot containing the character, side by side, printed or on a second monitor. Looking at the strip makes drift obvious in a way that watching shot by shot does not.

The resemblance problem

A generated face can resemble a real person by accident. Models learn from enormous numbers of real images, and a prompt describing a particular age, ethnicity, hairstyle and build may produce a face strikingly like an actual person. If that person is well known, the resemblance may be recognised by audiences and may create legal risk under laws on publicity, likeness and digital replicas, which vary by jurisdiction and are discussed in Chapter 19. If the person is not well known, the risk is lower but the ethical question remains.

Practical steps to manage it:

- **Never prompt with real names** for characters, and never use photographs of real people as references without written consent for the specific use.
- **Run a resemblance check** on principal synthetic characters before production: show the reference set to several people unconnected to the project and ask whether the character reminds them of anyone.

For commercial work, ask your legal adviser what further checks they expect.

- **Design away from resemblance.** If a character resembles someone, change features until it does not. It is cheaper now than later.
- **Record the check** in the character bible.

Casting real people for synthetic use

When a real person's likeness or voice will be used to create or drive a synthetic character, or a real performer will be digitally altered, consent becomes the central issue.

In the United States, the SAG-AFTRA agreements with the major studios set specific rules, which Chapter 19 describes in detail. In summary, the 2023 TV/Theatrical agreement required clear and conspicuous consent, separately signed, with a reasonably specific description of the intended use, before creating or using a digital replica of a performer, with advance notice and compensation. The 2026 agreement, ratified in June 2026, added further terms concerning fully synthetic performers. California law since January 2025 has made certain contract terms for digital replicas unenforceable unless the uses are reasonably specifically described and the performer was represented. Tennessee's ELVIS Act, in force since July 2024, protects voice and likeness against unauthorised simulation.

Even outside those rules, a fair practice for any studio is:

- **Specific consent**: a written agreement naming the project, the uses, the media, the duration and the territory. No blanket consent "for any purpose."
- **Separate signature** for the replica or alteration terms, so they are not buried in a general contract.
- **Fair payment** for the creation of the replica and for its use.
- **Approval rights** for the performer over uses that differ from what was described.
- **Data handling**: where the scans, recordings and trained adapters are stored, who can access them, and when they will be deleted.

- **Death and incapacity**: what happens to consent if the performer dies or cannot give approval, and who can decide on their behalf.

Kestrel Lane uses a consent template, adapted from Appendix A, for every performer whose likeness or voice is captured for anything beyond conventional filming. For *The Salt Road*, the actor playing the keeper consented in writing to a single, specific transformation: relighting of his performance in three shots to match generated lightning. He was shown the result before it was locked. That level of care costs very little and avoids a great deal.

Synthetic performers

Finally, a word on characters presented as performers in their own right: synthetic "actors" with names, social media accounts and careers. When one such character drew wide attention in September 2025, SAG-AFTRA issued a statement that creativity should remain human-centred, that the character was not an actor but a computer-generated character trained on the work of many professional performers without permission or compensation, and that producers signatory to its agreements must meet notice and bargaining obligations before using synthetic performers. The 2026 agreement made those obligations more specific.

A studio does not need to take a position on whether synthetic performers are good or bad to recognise the practical point: in film and television covered by union agreements, using one is a bargaining matter, and audiences, festivals and distributors are paying attention. If your project centres a synthetic performer, decide early how you will disclose it, what agreements apply, and how you will answer questions about where its appearance and performance came from.

The next part of the book moves from pre-production to production, beginning with the core skill of directing a video generator.

PART III

Production

CHAPTER 7

Directing the Generator

Directing a video model is a craft. It shares much with directing a camera crew and something with directing actors, and it has habits of its own. This chapter covers the core of that craft: writing shot descriptions a model can follow, choosing between text, image and reference inputs, using first and last frames and keyframes, extending and joining shots, planning coverage and selecting takes. It ends with a dated snapshot of the model landscape in November 2026 and advice on choosing among models without becoming dependent on any one of them.

The shot description

A prompt for a video shot is not a wish. It is a description of what a camera sees for a few seconds. The best prompts read like a precise line from a shooting script combined with a note from a director of photography.

A useful order, which puts the most important information first, is:

1. **Shot size and angle.** "Medium close-up, eye level." "Extreme wide shot from a high angle."
2. **Subject.** Who or what is in frame, described consistently with the character bible or reference set.
3. **Action.** What the subject does, in one or two simple verbs. "Turns from the window and looks toward the door."
4. **Camera movement.** "Static." "Slow push in." "Handheld, following from behind."
5. **Setting.** Where, described consistently with the location reference.
6. **Light.** Direction, quality, colour and source. "Cold daylight from a window on the left; warm lamp light behind her."

7. **Lens and texture.** “35mm, shallow depth of field.” Add texture only if you are not adding it in finishing.
8. **Sound**, if the model generates audio. “Wind against glass; a distant foghorn.”

Keep it to one shot and one main action. A prompt that describes a sequence of events invites the model to compress or skip them. If a shot needs two actions, consider whether it should be two shots.

Avoid words that describe evaluation rather than content. “Stunning”, “cinematic”, “masterpiece” and “award-winning” add little and can push results toward a generic gloss. Describe what makes the image stunning instead: the light, the composition, the moment.

Be consistent. Reuse the exact same wording for characters, locations and look across all shots in a sequence. Small variations in wording produce small variations in result, and those accumulate into continuity errors. The reusable prompt blocks from Chapters 5 and 6 exist for this purpose.

Choosing the input

Video models accept several kinds of input, as Chapter 3 explained. The choice of input is the most important directing decision for each shot.

Text-to-video gives the model the most freedom. Use it for exploration, for atmospheric shots where composition matters less, and for shots where you want to be surprised. Do not use it for shots that must match a storyboard or a neighbouring shot.

Image-to-video starts from a frame you have made and approved. It is the default for narrative work, because it fixes composition, character and look at the first frame. The prompt then needs to describe only motion, camera and sound.

First and last frame control gives the model both ends of the shot and asks it to fill the motion between. It is powerful for transitions, for shots that must end on a precise composition so they cut to the next, and for controlling where a camera move lands. Choose end frames that are physically reachable from the start frame in the duration; if they are too different, the model will invent an unconvincing transformation.

Keyframes extend the idea to several images along the shot. Some tools support them directly. They let you choreograph more complex motion, at the cost of more preparation.

Reference-to-video supplies subjects without fixing composition. Use it when you need the model to stage a shot freely but keep characters, props or locations consistent. It combines well with image input: a start frame for composition, references for identity.

Video-to-video transforms existing footage. Use it when you have a performance, a previs animation or a rough shot whose motion you want to keep while changing its look. Chapter 8 covers it in more detail.

Duration and the usable middle

Generated shots often have a weaker start and end. In the first moments, motion is beginning and the model is settling; at the end, motion may slow, freeze or drift. The usable part is often the middle.

Plan for this. If a shot needs four seconds on screen, generate six or eight and use the middle. If the model offers durations in fixed steps, choose the step above what you need. Many models of late 2026 allowed duration to be chosen in one-second steps within a range, which helps.

When a shot must be longer than a model's maximum, there are three options, in order of preference:

1. **Cut it.** Most long shots in conventional films are long for a reason. If yours is long only because the action takes time, a cut to a reaction or a detail is usually better.
2. **Join shots.** Generate two shots where the first frame of the second is the last frame, or nearly the last frame, of the first. Use first-frame control. Match colour and motion at the join, and consider hiding it with a cut on action or a brief cutaway.
3. **Extend.** Many tools can extend a clip by generating a continuation. Each extension risks drift in identity, light and colour. Inspect every join carefully.

Coverage and selection

The shooting ratio, the footage shot divided by the footage used, is an old concept that matters again. In generated production, every attempt costs money and review time. Planning coverage controls the ratio.

For each scene, Kestrel Lane plans:

- the **master** shots from the storyboard;
- one **alternative angle** for each important moment;
- several **cutaways**: details, hands, objects, landscape, reactions;
- a **safety** version of any risky shot, simplified so that it is very likely to work.

Then, for each planned shot, they generate a small batch of variations, review them, and either select, refine or redesign.

Select when one variation does the job, even if it is not perfect. Perfect shots are rare in any form of filmmaking.

Refine when a variation is close: adjust the prompt, the input frame or a setting, and generate again from the same seed if the tool allows it, so that only the thing you changed varies.

Redesign when several rounds of refinement fail. Break the shot into simpler shots, change the angle to hide the difficult part, move it to a different method (shot, captured or composited), or cut it. The single most expensive habit in generated production is repeating a failing shot in the hope that the next attempt will work.

Keep the selected take and one or two runners-up. Delete or archive the rest, but keep their records in the generation log. You will be surprised how often a rejected take solves a problem later in the edit.

Writing the generation log

Every generated shot that might reach the film should have a log entry. A template is in Appendix A. At minimum, record:

- the shot number;
- the date and time;
- the service or software, model name and version;

- the inputs: prompt text, input frames, references and their versions, any adapters;
- the settings: seed, duration, resolution, aspect ratio and any controls;
- the output file name and its checksum;
- the reviewer's decision and the reason.

For a small studio this feels like overhead. It repays itself in three ways. It lets you regenerate or repair a shot months later. It supports copyright registration, where you may need to describe the human contribution to AI-assisted material. And it answers the questions distributors, insurers and festivals increasingly ask about how a film was made. Chapter 16 shows how to automate most of it.

The model landscape, November 2026

What follows is a snapshot. It will date quickly. It is included to illustrate the kinds of capabilities a director could choose among at the time of writing, not to recommend any model.

As of November 2026, the leading commercial video models came from a mix of American and Chinese companies and from both large technology firms and specialist start-ups. A third-party comparison published on 2 November 2026, aggregating rankings from public blind-preference leaderboards, placed ByteDance's Seedance 2.5, MiniMax's H3, Alibaba's Wan 3.0 and Google's Gemini Omni at the top, with Black Forest Labs, xAI, Kuaishou's Kling, ShengShu's Vidu and Runway among the other leading providers. Rankings of this kind move month by month and depend on the prompts tested.

Several capabilities had become common among leading models:

- clips of fifteen to thirty seconds in a single generation, with some models extending further;
- synchronised audio, including dialogue, effects and music;
- reference inputs for characters, objects and locations, in some cases numbering in the tens;
- first-frame and last-frame control;

- native or upscaled output at 1080p or 4K; Kling, for example, announced native 4K output in mid-2026 and 4K HDR with Kling 4.0 in October 2026;
- conversational editing, in which a follow-up instruction changes a generated clip rather than regenerating it from scratch. Google described multi-turn editing for Gemini Omni 1.2 Flash, launched on 5 November 2026 with clips of three to fifteen seconds and a published price of US$0.07 per second at 720p.

Several companies offered open-weight models, which a studio can run on its own hardware and adapt. Lightricks released LTX-2, an audio-video model, with open weights in January 2026; Alibaba's Wan line had open-weight versions; and the third-party comparison listed MiniMax's H3 as the strongest open-weight model at the time.

The landscape also showed how quickly a leading model can disappear. OpenAI closed its Sora app and website in April 2026 and its API in September 2026. Productions and partners that had built on it had to move.

Choosing among models

Given that churn, how should a director choose?

Test on your own shots. Leaderboards average over many prompts. Your film has particular needs: a character, a look, a kind of motion. Keep a small **test suite** of five to ten representative shots from your current project, with their input frames and references, and run new models against it. Judge the results against your character bible and look document, not against general impressions.

Match model to shot. Different models are better at different things: some at human motion, some at landscapes, some at stylised animation, some at dialogue. Many productions use two or three models, chosen shot by shot. Record which, in the log.

Read the terms. Each service has terms about ownership of outputs, permitted commercial use, data retention, training on your inputs and content restrictions. They differ, and they change. For commercial work,

read them and keep a dated copy with the project. Chapter 18 explains why.

Keep assets model-independent. Your key frames, references, look documents, character bibles and previs can drive any model. Trained adapters are tied to a specific base model. Prompts often need light rewriting for a new model. Store everything so that switching is a matter of days, not a rebuild.

Have a second source. For any production on a schedule, identify at least one alternative model for each kind of shot you need, and test it before you need it.

A worked example: the storm arrives

To see these practices together, follow one short sequence from *The Salt Road*: the moment the storm reaches the coast. In the production draft it is five shots, about twenty seconds on screen.

1. **Extreme wide shot**, high angle: the headland and lighthouse, a wall of dark cloud advancing from the sea. Generated.
2. **Medium shot**: the surveyor's car on the coast road, wipers beginning, the first heavy drops. Composited: the car and actor shot in a garage, the road and storm generated.
3. **Close-up**: the keeper at the lamp-room window, looking out. Shot, later relit.
4. **Insert**: a gull struggling against the wind above the cliff. Generated.
5. **Wide shot**: the first wave breaking over the sea wall beneath the lighthouse. Generated.

For shot 1, the director composed a key frame with an image model from the look document's reusable block, the lighthouse exterior reference set and a simple blocking render of the headland made from the captured splat of the lighthouse top. The key frame went to image-to-video with a short prompt: "The cloud wall advances slowly toward the headland from the right. Static camera. Rain begins to sweep across the sea. Low wind, distant thunder." The first batch of four drifted toward a blue sea; the reusable block's "the sea is never blue" line had been omitted from

the prompt by mistake. Restored, the second batch produced two usable takes. The middle five seconds of the better one were selected.

For shot 4, the gull, the first three batches produced birds with wrong wing motion or extra wings. Rather than repeat, the director redesigned the shot: a smaller bird, farther from camera, partly hidden by rain, with less screen time. The first batch of the redesigned shot worked. On screen, nobody misses the close gull.

Shot 5 needed to cut on the wave's impact into the next scene. It was generated with first-and-last-frame control: a key frame of the calm sea wall and a key frame of spray at its height, both composed from the same blocking render. The model filled the wave between them. Two attempts were needed, because in the first the wave broke too early and the remaining frames held on spray.

The log for this sequence records twenty-six generations for five shots, including drafts. Most of the cost came from shots 4 and 5. Most of the quality came from the key frames, which took longer to make than all the generation combined.

The director's role

It would be easy to read this chapter as a set of technical procedures. It is better read as a description of what a director has always done: deciding what each shot is for, describing it precisely to the people or tools who will make it, judging the results honestly, changing the plan when the plan does not work, and keeping track of what has been made.

The model is fast, tireless and literal. It does not know what your film is about. Every choice that makes a shot belong in your film, and not in someone else's, is yours. The next chapter covers the tools that give you finer control over two of those choices: where the camera goes, and how things move.

CHAPTER 8

Camera, Motion and Control

The difference between a generated clip and a directed shot is control. A clip shows you what the model imagined. A shot shows the audience what you decided they should see, from where, moving how, for how long. Through 2024 and 2025, the most important advances in video generation for filmmakers were not in image quality but in control: over the camera, over motion, over structure and over existing footage. This chapter covers those controls, how they work, and how to combine them into a dependable method for getting the shot you planned.

The language of camera movement

Every camera move has a purpose. A push in draws attention and raises tension. A pull out reveals context or isolates a character. A pan or tilt follows action or connects two things in a space. A tracking move travels with a subject. A crane rises or descends to change the relationship between a character and their world. A handheld camera adds presence and instability. A static frame lets the action play.

Video models learned these moves from their training data and generally understand the common terms: dolly in and out, pan, tilt, truck, pedestal, crane, orbit, handheld, static. Using them in a prompt works reasonably well for simple moves. It works less well for combined moves, for precise speed and distance, and for moves that must land on an exact composition. That is where explicit controls come in.

Camera control

Research in 2024 and 2025 showed how to make a video model follow a specified camera path. MotionCtrl, presented at SIGGRAPH 2024, controlled camera motion and object motion separately. CameraCtrl, presented at ICLR 2025, added a camera-pose module to existing video

models so they could follow a precise trajectory, and a successor presented at ICCV 2025 extended the idea to longer, dynamic scene exploration. The common principle is that the camera's position and orientation for each frame are given to the model as an additional input, alongside the text and images.

Commercial tools expose camera control in several forms:

- **Named presets**: a menu of standard moves such as push in, orbit left or crane up, sometimes with an intensity setting.
- **Direction and speed controls**: sliders for horizontal, vertical, zoom, pan, tilt and roll movement.
- **Explicit paths**: a camera path drawn or keyed in a 3D view, or imported from a 3D application.
- **Motion from reference video**: a clip whose camera move you want copied.

As a rule, the more explicit the control, the more reliable the result, and the more preparation it needs. For most narrative shots, a preset or simple direction controls combined with a good start frame and, ideally, an end frame, is enough. For shots that must match a previs, match a photographed plate or land precisely, use explicit paths.

Using previs to drive the camera

The most dependable way to control camera movement is to design it in a 3D environment and pass it to generation. The workflow has four steps.

1. **Block the scene in 3D.** Use simple geometry, mannequins or a Gaussian splat of a real location (Chapter 9). Put the camera where you want it and animate its move. Time it to the edit.
2. **Render control passes.** From the 3D scene, render the camera's view as a simple image sequence, and, where the generation tool accepts them, as depth maps, edge maps or pose skeletons for characters.
3. **Generate with structural conditioning.** Feed the control passes to a video model or workflow that accepts structural inputs, together with the look, references and prompt. The model renders the shot in the film's style while following the composition and camera of the previs.

4. **Check against the previs.** Compare the generated shot to the previs frame by frame. Accept small differences; reject changes of composition that break the cut.

This approach is most accessible in node-based tools and open-weight models, where depth, pose and camera inputs can be wired together, as Chapter 16 describes. Some commercial services offer parts of it, such as importing a rough 3D render as a guide. It takes more setup than writing a prompt, but it gives a level of control comparable to conventional animation, and the 3D blocking remains useful if a model changes.

For *The Salt Road*, Kestrel Lane used this method for the storm's single most important shot: a long, descending crane move from above the lighthouse, through the rain, to the keeper at the gallery rail. They blocked it in Blender against a splat of the real lighthouse top, rendered depth and a camera sequence, and generated the final shot from those passes with the storm look and the keeper's references. It took two days. Earlier attempts from text and presets alone had taken three days without producing a usable take.

Motion control and motion transfer

Camera control decides where the camera goes. Motion control decides how things in the frame move.

Motion brushes and trajectories. Some tools let you paint regions of a start frame and assign each a direction of motion, or draw a path for an object to follow. They are useful for simple, clear movements: clouds drifting one way, a boat crossing the frame, a door opening.

Pose and skeleton conditioning. A sequence of body poses, extracted from a recorded performance or animated by hand, can drive the movement of a generated character. The model generates a character with the right appearance moving through the given poses. This is the basis of most character motion control.

Motion transfer from video. Several tools take a reference video of a person moving and transfer the movement, and often the facial performance, to a character defined by an image or reference set. Runway released Act-Two for this purpose in July 2025, transferring

head, face, body and hand motion from a driving performance to a character. Kling announced an upgraded motion-control capability with its 3.0 model in March 2026, which it described as professional motion capture with high facial consistency. Chapter 11 discusses performance capture in more depth.

Motion transfer changes the economics of character work for small studios. An actor can perform a scene in a plain room in front of a phone, and that performance can drive a generated character in a generated world. The actor's timing, weight, gesture and expression come through, and these are exactly the qualities text prompts struggle to produce. It also raises the consent questions covered in Chapter 6: the performance is a real person's work, and the agreement should say how it will be used.

Video-to-video: transforming footage

Video-to-video models take an existing clip as input and produce a transformed version. They are as important for production as for post-production, because they turn any rough footage into a controllable starting point.

Runway's Aleph, announced in July 2025, described a wide range of such operations: adding, removing and changing objects, changing environments, weather and lighting, applying styles, and generating new camera angles or the next shot from an existing one. Runway later announced Aleph 2.0 alongside an editing environment. Other providers released comparable editing capabilities, and several leading models of 2026 offered editing modes alongside generation.

Typical production uses include:

- **Restyling previs or rough animation** into the film's look while keeping motion and composition.
- **Turning a rough shoot into a finished shot**: a performer shot in a garage becomes a performer on a ship's deck.
- **Changing conditions**: day to dusk, dry to rain, summer to winter, for continuity or story.

- **Alternative coverage**: generating a reverse angle or a closer shot from an existing take. Treat these as useful but risky: a generated angle on a real performance invents parts of the performance that were not captured, which may matter for consent and for accuracy.

The quality of video-to-video output depends heavily on the input. Clean, well-lit, well-composed rough footage transforms far better than murky footage. If you plan to transform a shot, shoot it with that in mind: clear separation between subject and background, steady exposure, and the composition you want in the final.

Consent and video-to-video

One caution deserves its own section. Video-to-video tools make it trivially easy to change a real performer's appearance, surroundings, age, clothing and even apparent expression. When the input footage shows a real person, every such transformation is a potential alteration of their performance or likeness. Routine changes, such as grading, cleanup and modest environmental changes behind them, are part of normal post-production. Changes to the person themselves, such as their face, body, apparent age, expression or speech, are the kind of alteration that the agreements and laws in Chapter 19 address, and that good practice requires consent for. Before running a performer's footage through a transformation, check what was agreed, and record the transformation in the shot's log.

Keyframes and in-betweens

Animators have always worked with key poses and in-betweens: the animator draws the important moments and the in-betweens fill the motion between them. First-frame, last-frame and keyframe controls bring this method to generation.

A practical keyframe workflow:

1. Design the key moments of the shot as still frames, using the image model, references and structural guides so that composition, character and light are exactly right.

2. Check that the keys are physically compatible: positions reachable in the time, light consistent, nothing appearing or disappearing without cause.
3. Generate the motion between them with first-and-last-frame or keyframe control.
4. Review the in-betweens for invented action, which models sometimes add to fill time.

This is the most dependable way to get a shot that both begins and ends exactly where the edit needs it. It also gives you strong human-authored material, in the form of the designed keyframes, which matters for the authorship questions in Chapter 18.

Combining controls

The real power comes from combining controls. A shot might use:

- a designed start frame and end frame for composition;
- reference images and a trained character adapter for identity;
- a camera path from previs;
- a pose sequence from a recorded performance;
- a short, precise prompt for light, atmosphere and sound.

Not every tool accepts every input, and not every combination works well together; controls can conflict, producing artefacts where the model cannot satisfy them all. Start with the fewest controls that give you the shot, and add more only when a specific problem calls for it. Record every control in the generation log.

Physical camera thinking

Even with every control available, generated shots look most convincing when they obey the logic of a physical camera. A few principles help:

Choose a lens and stick to it. A shot described as a 35mm wide shot that then behaves like a telephoto close-up will feel wrong. Describe the lens, and keep camera moves consistent with it.

Respect the camera's weight. Real cameras accelerate and decelerate; they do not jump to full speed. Real handheld cameras sway with a body's rhythm. Explicit paths should ease in and out.

Motivate the move. A camera that moves for no reason draws attention to itself. Generated shots often include unmotivated drift. If the move is not doing something, ask for a static camera.

Keep the horizon honest. Models sometimes tilt or bend horizons during moves. Check the horizon in every landscape shot.

Match coverage to geography. When cutting between angles in a scene, keep the screen direction consistent: if a character exits frame left, they should enter the next shot from the right. Models know nothing about your previous shot's screen direction; you must specify it.

Covering a dialogue scene

Dialogue scenes test control more than any other kind, because they depend on consistent geography, matched eyelines and performances that play across cuts. Kestrel Lane avoided generating dialogue on screen for its principal characters, shooting those scenes with the actors. But some projects must generate dialogue scenes, and the approach that works best combines several controls.

First, establish the geography in a single wide **master** composed as a key frame: both characters, their positions, the room, the light. Every later shot in the scene is designed from this frame. Mark the line between the characters, sometimes called the axis or the 180-degree line, and keep every camera on one side of it.

Second, design the **singles** and **over-the-shoulder** shots as key frames derived from the master: same light direction, same background elements, eyelines pointing toward where the other character is. A simple 3D blocking of the room, with mannequins at the characters' positions, makes this precise; render each camera position from it as a structural guide.

Third, record the **performances**. An actor reading each character's lines, filmed in roughly the framing of each shot, drives the character through motion transfer or provides the audio that drives lip-sync. If the two

characters are played by different actors, record them reading together where possible, so the timing of the conversation is real.

Fourth, **generate shot by shot**, with each character's references or adapter, the key frame and the performance, and keep camera movement minimal. Dialogue scenes rarely need moving cameras, and static frames make matching easier.

Finally, **cut on the performance**. The editor chooses reactions and line readings first, and fixes visual mismatches with framing, flips where screen direction allows, and cutaways.

This is more work than generating a dialogue scene from a single prompt. It is also the difference between a scene that plays and one that looks like two unrelated clips of people talking.

When control fails

Some shots resist every control. A complex physical interaction, a crowd with specific choreography, a precise stunt. The sensible responses are those from Chapter 7: simplify, break into smaller shots, hide the difficulty with framing, or move to another method. Real capture, composited with generated elements, remains the most reliable way to get complex physical action. The next three chapters turn to capture: of real places as 3D scenes, of real performers on LED volumes, and of real faces and bodies driving digital characters.

CHAPTER 9

Capturing the Real: Photogrammetry, NeRFs and Gaussian Splats

The more a film depends on generated images, the more valuable real places become. A generated landscape is plausible; a captured one is true, down to the way light falls on a particular wall and the exact distance between two doors. Between 2020 and 2026, a sequence of techniques made it possible to capture real places and objects as three-dimensional scenes that can be viewed from new angles with photographic realism, using nothing more than a camera and a computer. By November 2026 the most important of them, Gaussian splatting, had native support in Blender, growing support in Unreal Engine, and a place in professional film, television and music video production. This chapter explains these techniques, how to capture well, and how captured scenes fit into a synthetic studio's pipeline.

Three generations of capture

Photogrammetry is the oldest. It takes many overlapping photographs of an object or place, finds matching points between them, works out where each camera was, and builds a textured 3D mesh. Photogrammetry has been used in film, games and heritage for decades. Its meshes work in any 3D software and any renderer, and can be lit like any other geometry. It struggles with shiny, transparent and thin surfaces, with fine structures such as hair, foliage and fences, and with view-dependent effects such as reflections.

Neural radiance fields, or NeRFs, were introduced in a 2020 paper by Mildenhall and colleagues. Instead of a mesh, a NeRF represents a scene as a neural network that, given a point in space and a viewing direction,

returns colour and density. Rendering a view means sampling the network along rays from the camera. NeRFs captured view-dependent effects and fine detail that photogrammetry missed, and they produced remarkable new views. They were slow to train and, in their original form, slow to render, which limited their use in real-time production.

3D Gaussian splatting, introduced by Kerbl, Kopanas, Leimkühler and Drettakis at SIGGRAPH 2023, represents a scene as millions of small, semi-transparent, coloured ellipsoids, called Gaussians, each with a position, shape, orientation, opacity and a colour that can vary with viewing direction. Training adjusts the Gaussians until rendered views match the input photographs. Rendering is fast, because the Gaussians are simply projected onto the screen and blended, which modern graphics hardware does very efficiently. Splats combine much of the photorealism of NeRFs with real-time rendering, and that combination is why they spread so quickly.

Why capture matters to a synthetic studio

Captured scenes do several jobs that generation alone cannot.

Geographic truth. A captured location keeps its real layout. A camera can move through it and see what is really there. This matters for films set in real places and for any sequence where spatial continuity matters.

Real-time previs and blocking. A splat of a location can be loaded into a 3D application or game engine and used for blocking and camera planning before a shoot, as Kestrel Lane did with the lighthouse lamp room.

Plates without a shoot. A virtual camera moving through a splat produces plates that can be composited with performers shot elsewhere, or used as structural guides for generation.

Conditioning for generation. Depth and colour renders from a captured scene make excellent structural inputs for video-to-video or depth-conditioned generation, which can restyle the real place while keeping its geometry.

Backgrounds for LED volumes. Captured scenes can be displayed on LED walls in virtual production, as Chapter 10 describes.

Assets that last. A captured location is a durable, model-independent asset. If a generative model is withdrawn, the capture is still there.

Planning a capture

Good capture starts with permission. Capturing a private interior, a business, a heritage site or a recognisable artwork may require the owner's consent, and capturing people requires theirs. Some places restrict photography or drone use. Get written permission for the specific purpose, record it with the capture, and treat captured places with the same care you would treat a location for a conventional shoot.

Then plan the capture like a small shoot.

Decide what you need. A full 360-degree environment, or only the views the film needs? A whole building, or one room? Capturing only what you need saves time and produces better results, because effort concentrates on the parts that matter.

Choose the light. Captured scenes record the light present during capture. Soft, even, unchanging light, such as an overcast day or a well-lit interior with no direct sun, is easiest. Moving shadows, changing clouds and flickering lights cause errors. If you will relight the scene later, flatter light gives you more freedom. If the captured light is the look you want, capture at the right time of day and work quickly.

Choose the camera. Any good camera works. A camera with a fixed, sharp lens and manual exposure gives more consistent results than a phone's automatic modes. Many people capture with video, extracting frames afterwards; stills give higher quality per frame. Some productions use arrays of synchronised cameras to capture people or moving action, discussed under 4D capture below.

Lock exposure and white balance. Changes between frames confuse training. Set them manually and leave them.

Capturing

The essential rule is overlap. Each part of the scene should appear in many frames, from different positions. Move the camera through the

space rather than rotating on the spot; a camera that only turns in place gives no depth information.

For an object, circle it at two or three heights, keeping it in the centre of frame, with each frame overlapping the previous one heavily.

For an interior, move along the walls and through the centre, looking inward and outward, at several heights. Capture details from closer where you need detail.

For an exterior, walk the area in overlapping passes, as a surveyor would, at more than one height if you can.

Avoid motion blur, which ruins frames. Use a fast enough shutter speed, and move slowly. Avoid people and moving objects in frame where possible. Shiny and transparent surfaces, such as glass, polished metal and water, remain difficult for all methods; capture them anyway, but expect to fix or replace them.

Record the capture: date, time, location, camera and settings, number of frames, permissions, and anything unusual.

Training and cleaning

Training turns frames into a scene. Several kinds of tools existed in late 2026.

Desktop applications run training on your own computer and graphics card. Jawset's Postshot, for example, trained splats locally from images or video, exported to common formats, and offered integration with Unreal Engine and Adobe After Effects according to its documentation. Running locally keeps the data private, which matters for locations captured under confidentiality.

Open-source tools from the research community give the most control and support the latest methods, at the cost of more technical setup.

Cloud services and phone apps offer convenience for quick captures.

Whatever the tool, training a splat goes through recognisable steps: estimating camera positions from the frames (often called structure from motion), initialising Gaussians from the resulting sparse points, and optimising them against the frames. Check the camera estimation

first; most bad splats come from bad camera estimates, which come from insufficient overlap or inconsistent frames.

A trained splat usually needs cleaning. **Floaters**, stray Gaussians in empty space, appear where the training data was thin. Areas outside the region of interest are blurry. Editing tools let you crop, delete and select regions. Some tools can segment splats by object.

Using splats in production tools

Tool support for splats grew rapidly in 2025 and 2026, and November 2026 was a notable month. Blender 5.3, released on 10 November 2026, added native Gaussian splat support: importing common splat formats including PLY and SPZ, editing splats as point clouds with Geometry Nodes, and rendering them in Blender's Workbench, EEVEE and Cycles renderers, according to the release notes. For Unreal Engine, splats had been usable through third-party plugins, and reports in November 2026 described native support arriving in the engine. A newsletter on the subject that month described it as an exceptional few weeks for the technology.

The practical consequence is that a studio can now treat a splat much like any other 3D asset: place it in a scene, put a virtual camera in it, combine it with meshes, characters and effects, and render. Splats still have limits. They carry baked-in lighting, so relighting them convincingly is difficult, although research on relightable splats is active. Their file sizes can be large. Compositing them with conventional geometry requires care with depth and transparency. Their edges, especially near thin structures, can look soft.

A captured lamp room

Kestrel Lane's lamp-room capture illustrates the workflow. With the lighthouse owner's written permission, the producer captured the room on an overcast morning using a stills camera at fixed exposure, in about four hundred frames taken in overlapping passes at three heights. Training took an afternoon on their workstation. Cleaning removed floaters around the windows and simplified the view outside, which would later be replaced with the generated storm.

The splat then served three purposes. The director used it for blocking the one-day shoot, placing virtual cameras in Blender to plan coverage, which saved hours on the day. It supplied six captured shots: slow virtual camera moves across the great lens and the brass machinery, rendered from the splat, with the storm composited outside the windows. And it supplied depth and colour passes for two generated shots, in which a video model restyled the real room into its appearance during a power cut, lit only by lightning, while the geometry stayed true.

4D capture: splats in motion

A static splat captures a place. Capturing people and action in three dimensions needs time as well: so-called 4D capture. The common approach uses a synchronised array of cameras surrounding the action, each recording video, with splats trained for each moment in time or with methods that model how Gaussians move.

4D capture was moving from research into production in 2025 and 2026. One capture company described a production using forty-eight cameras together with an LED wall. Netflix's visual effects studio Eyeline was reported in November 2026 to be working on improving Gaussian splats for volumetric capture of performances, and a Netflix internship listing described investigating splats as a possible future streaming format, although that work appeared to be at an early stage.

For a small studio, 4D capture remains expensive and specialised. Its significance is directional. It points toward performances that can be re-photographed after the fact from any angle, which would change cinematography, editing and visual effects; and toward audiences who might one day move freely within a scene. Chapter 21 returns to this.

Combining capture and generation

The most interesting work in 2026 combined capture and generation rather than choosing between them. Patterns that worked included:

- **Captured geometry, generated look.** Render depth and colour from a splat; restyle with a video model under structural conditioning. The space stays true; the time, weather, period or style changes.

- **Generated extension of captured space.** Use the real captured room as the foreground; generate what lies beyond its windows or doors.
- **Captured background, shot performer.** Composite a performer shot against a plain background into a virtual camera move through a splat.
- **Generated environments to 3D.** Some research systems and products generate 3D scenes, including splats, from images or text. These are useful for previs and backgrounds. As of late 2026 they were less reliable than real capture for detail and accuracy.

When a capture goes wrong

Captures fail in recognisable ways, and most failures can be traced to the session rather than the software.

Camera estimation fails or splits. The software cannot work out where some frames were taken, or produces two disconnected groups. Cause: too little overlap, or a sudden change of view. Fix: recapture the gap with more frames, or add frames bridging the two groups.

Blurry or smeared regions. Cause: motion blur, focus errors, or too few views of the area. Fix: remove bad frames before training; recapture the area with more, sharper frames.

Floaters and fog. Cause: thin coverage of empty space, reflections, or moving elements. Fix: clean in an editor; mask out moving objects; recapture reflective areas from more angles.

Holes behind objects. Cause: areas never seen by the camera. Fix: capture from more positions, including low and high; accept that some areas will need to be hidden from the virtual camera.

Wrong scale. Captures have no inherent scale. Fix: include an object of known size, or measure a distance on site, and scale the result.

Colour shifts across the scene. Cause: automatic exposure or white balance, or changing light. Fix: lock settings; capture faster; correct in training where tools allow.

Keep notes of failures and fixes. Capture is a skill, and a studio's capture checklist improves with every session.

Capturing for generation

When a capture will be used mainly as conditioning for generation, rather than rendered directly, the priorities shift a little. Geometry matters more than surface detail, because the generative model will supply much of the look; depth and edge passes need to be clean and stable. Coverage of the specific camera paths you plan matters more than complete coverage of the space. And flat, even light is preferable, because baked-in shadows in the colour pass can confuse a model asked to apply a different light.

A useful habit is to capture twice in the same session if time allows: a careful, complete pass for direct rendering, and a faster pass along the planned camera routes, which can be trained separately for previs and conditioning. The two serve different purposes, and the second is often enough on its own for generated shots.

What to keep

Treat captures as primary assets. Keep the original frames, the camera estimates, the trained scene, the cleaned scene, the permissions and the capture record together. Frames are the most important: training methods will improve, and you can retrain from good frames for years. A cleaned splat is useful now; the frames are useful indefinitely.

The next chapter takes captured and generated worlds onto the stage, where they surround real performers on walls of LED light.

CHAPTER 10

Virtual Production and the LED Volume

For most of the history of visual effects, performers acted against blue or green screens and imagined the world that would later be composited behind them. The director, the cinematographer and the actors worked with a promise. In 2019, the first season of *The Mandalorian* showed a different approach at scale: performers on a stage surrounded by a curved wall and ceiling of LED panels displaying photorealistic environments rendered in real time by a game engine, with the view on the wall shifting in perspective as the camera moved. The performers could see the world. The camera captured the final image, including the environment's light on their faces and reflections in their costumes, in a single pass. This chapter describes that approach, called virtual production or in-camera visual effects, how it works, what it costs, and how generative AI and captured scenes are changing what goes on the wall.

What virtual production means

"Virtual production" is used broadly to mean any production process in which real-time computer graphics are used during shooting or planning. It includes virtual scouting in a 3D model of a set, real-time previs, motion-capture stages where a director sees animated characters live, simulcam setups that overlay a virtual scene on the camera's view, and LED-volume shooting. This chapter focuses on the last, which is often called **in-camera visual effects**, or ICVFX, because the effect is captured by the camera rather than added afterwards.

How an LED volume works

According to Industrial Light & Magic and Epic Games, who described the system publicly, the stage used for *The Mandalorian*, which ILM called StageCraft, combined several components. An LED wall and ceiling surrounded the performance area. The Unreal Engine game engine rendered the virtual environment. Camera tracking reported the position and lens of the physical camera to the engine in real time. A cluster of computers, using Unreal's nDisplay system, drove the panels. Operators adjusted lighting, effects and animation in the virtual scene live during shooting. ILM reported that more than half of the first season was shot this way.

The key idea is the **frustum**: the region of the LED wall that the camera actually sees. Inside the frustum, the engine renders the environment from the exact viewpoint of the tracked camera, so that when the camera moves, the background shifts with correct parallax, exactly as a real distant landscape would. Outside the frustum, the wall displays a lower-detail version of the environment, which the camera does not see but which lights the performers and set, and appears in reflections.

That is the source of most of the advantages of LED volumes:

- **Interactive light.** The environment lights the performers. A sunset on the wall puts sunset light on their faces.
- **Reflections.** Shiny costumes, helmets, cars and glass reflect the environment naturally, which is very hard to achieve with green screen.
- **Performance.** Actors see and react to the world rather than imagining it.
- **Decisions on the day.** The director and cinematographer see the final image through the lens and can change the time of day, move a mountain or adjust the sky between takes.
- **Fewer composites.** Many shots need little or no post-production compositing.

And of most of the challenges:

- **Content must be ready before the shoot.** The environment has to be built, optimised to run in real time and approved before the stage day. Virtual production moves visual effects work from post-production to pre-production.
- **Cost.** Stage time, crew, engine artists and the computers to drive the wall are expensive. Rental stages have made volumes more accessible, but they remain a significant budget line.
- **Technical limits.** Moiré patterns can appear when the camera focuses on the wall. Colour from LED panels differs from natural light and needs calibration. Latency between camera movement and wall update can cause the background to lag on fast moves. Very wide shots may show the edges of the volume.
- **Physical integration.** The join between the physical set on the stage floor and the virtual environment on the wall must be carefully designed and lit.

A smaller volume

Volumes range from the large, curved stages used by major series to modest flat walls a few metres wide. Smaller productions often use a single flat LED wall, sometimes with additional panels or conventional lights for fill, combined with a tracked camera. This is enough for many shots: a car interior with a moving background, a window looking onto a virtual landscape, a character against a distant environment in medium shot.

For even smaller budgets, the same principles apply with a large television or projector for backgrounds behind windows, and conventional lighting designed to match the virtual environment. The essential practice, which is to have the world visible and lighting the performers during the shoot, scales down.

Kestrel Lane did not have a volume budget for *The Salt Road*. They rented time on a small LED wall at a local facility for a half-day to shoot the two lighthouse-window close-ups, displaying the generated storm outside the

window so that the lightning lit the keeper's face on camera. Every other storm interaction was composited.

What goes on the wall

The content displayed on a volume is the heart of the matter, and this is where generative AI and capture have had the biggest effect.

Traditional real-time environments are built by 3D artists in a game engine from modelled, textured and lit assets. They give complete control, real-time parallax and the ability to change anything, at the cost of significant artist time.

Plates are video recorded on location, sometimes with arrays of cameras, played back on the wall. They are photorealistic and relatively cheap but have fixed perspective, so they suit shots where the camera does not move much relative to a distant background, such as driving shots.

Captured scenes such as Gaussian splats, described in Chapter 9, offer real-world realism with real-time parallax. A capture company described a production in 2026 that combined a forty-eight-camera array with an LED wall. Growing native support for splats in Unreal Engine, reported in November 2026, should make this route easier.

Generated environments can be used in several ways:

- **Generated imagery as textures and backdrops**: generated skies, distant landscapes and matte paintings placed on far layers of a real-time environment, where limited parallax is needed.
- **Generated 2.5D and 3D environments**: generated images converted into layered or 3D scenes, increasingly with tools that generate splats or meshes from images.
- **Generated video on the wall**: generated clips played back like plates, for moving skies, water, weather and crowds in the distance.
- **World models**: interactive generated environments that respond to camera movement in real time are, in principle, an ideal fit for a volume. As of November 2026, they were research systems with limited resolution, latency and consistency, and were not ready for production walls. Chapter 21 discusses the direction.

Planning a volume shoot

Because content must be ready before the stage day, a volume shoot is planned like a visual effects shoot that happens in reverse.

Previs everything. Every shot is previsualised with the environment, the camera position and the lens, to confirm what the camera will see, where the frustum falls and where the edges of the volume are.

Build content to the shots. Detail is spent where the camera looks. Distant areas can be simpler. Generated elements are well suited to areas that need realism but not interactivity.

Hold a tech scout. Load the environment on the actual wall before the shoot day. Check colour, brightness, moiré, frame rate and latency with the actual camera and lenses. Many problems show only on the real wall.

Design the floor. The physical set on the stage floor must blend into the virtual environment. Match materials, colours and lighting, and design the join line so the camera does not see it, or so it is hidden by set pieces.

Plan the light. The wall lights the scene, but it rarely does everything. Conventional lights add key light, hard sunlight and accents. Lighting is designed together with the environment.

Prepare to change. Have variants ready: different times of day, weather or set dressing that operators can switch between takes.

Colour on the volume

Colour management is one of the hardest technical issues in LED production. LED panels emit light in narrow spectral bands, unlike sunlight or most film lights, so colours on camera can differ from what the eye sees, and skin tones lit by the wall can shift. The camera, the wall, the engine and the grading pipeline must be calibrated together. Productions typically involve the colourist and the camera team in content preparation and the tech scout, and work in a consistent colour pipeline, such as the ACES system described in Chapter 15, from the engine through to finishing.

Generated content adds a further step. Generated images and clips are usually delivered in display-referred formats with standard dynamic range, while volume content is ideally scene-referred with high dynamic range so that the wall can produce bright, realistic light. Converting generated content for the wall needs care to avoid flat, dim or banded results. Where possible, generate at higher bit depth or use upscaling and inverse tone mapping tools designed for this conversion, and check on the wall.

A half-day on a small wall

Kestrel Lane's half-day on a rented LED wall, for two close-ups of the keeper at the lamp-room window, illustrates how small volume work is planned.

Two weeks before, the producer sent the facility the generated storm clips intended for the window, a description of the shots, the camera and lenses, and the frame rate. The facility's technician converted the clips for playback and flagged two that would band visibly on the wall; they were regenerated at higher quality and run through a light debanding pass.

A window frame and a section of wall, built cheaply from the lamp-room references, stood a little in front of the LED panels. Because the camera saw the wall only through the window, at a distance and slightly out of focus, camera tracking and real-time parallax were not needed: the storm played back as a plate. That kept the cost down. The wall's job was mainly light: the flashes of generated lightning lit the actor's face from the right direction at the right moments.

The technician calibrated the wall to the camera at the start of the session, and the director of photography added a soft key and a practical lamp. The actor could see the storm and react to each flash. In two hours they had four good takes of each close-up. The remaining time was used to shoot clean reference passes of the window with no actor, for later compositing.

The relighting described in Chapter 13 was still needed for three other shots of the keeper, filmed on the main shoot day without the wall. Comparing those shots with the wall shots showed the difference clearly:

the wall shots needed almost no compositing, and their light felt entirely natural.

Preparing generated content for the wall

When generated imagery or video is destined for an LED wall, it should be prepared as carefully as any other volume content. A short checklist, which the facility technician and the colourist should agree:

- **Resolution and frame rate** suited to the wall and to the camera's shutter and frame rate, to avoid judder and flicker.
- **Bit depth and banding**: generate at the highest quality available, upscale and process in high bit depth, and check gradients on the actual wall.
- **Brightness and dynamic range**: convert standard-dynamic-range material into the wall's working range carefully, so skies and highlights are bright enough to light the scene without clipping or flattening.
- **Seamless loops**: background clips often need to loop during long takes; generate or edit them to loop without a visible join.
- **Consistency across variants**: if operators will switch between times of day or weather states, check that the variants match in perspective and layout.
- **Provenance**: log generated wall content like any other generated material; it is part of the final image.

Virtual production without a volume

The principles of virtual production, which are deciding the world before the shoot and seeing it during the shoot, can be applied without LED walls.

Virtual scouting. Explore a 3D environment, a splat of a real location or a generated world model on a screen or in a headset to plan shots, as described in Chapter 5.

Simulcam. Overlay a real-time view of the virtual environment on the camera's monitor during a green-screen shoot, so the director and

camera operator can frame against the final world. This is widely used for visual effects-heavy shooting.

Real-time compositing. Composite performers against virtual environments live, for monitoring or, in some broadcast uses, for final output.

Performance capture stages. Capture actors' movements and faces to drive digital characters in a real-time engine, with the director viewing the result live. Chapter 11 covers this.

Virtual production and generation: a convergence

There is a deeper connection between virtual production and generative video. Both are attempts to make the final image available earlier: to collapse the separation between shooting and visual effects. Virtual production does it by rendering the world in real time on set. Generation does it by producing images directly from descriptions and references.

The most interesting productions of 2025 and 2026 combined them. Real performers on a stage, lit by generated or captured environments on a wall, captured in camera, then extended or adjusted in post-production with video-to-video models. Or the reverse: generated previs and world-model exploration deciding the shots, captured scenes building the environment, and the shoot happening on a volume. The common thread is that the world is decided and visible before the camera rolls, and that real performances anchor the image.

What a small studio should take from this

Few small studios will build an LED volume. All of them can use the ideas:

- decide the world before shooting, in previs that the whole team can see;
- capture or generate environments early, and treat them as production assets;
- let the world light the performers whenever possible, even with a projector or a television behind a window, rather than adding everything in post;

- calibrate colour across the whole pipeline;
- rent volume time for the few shots where interactive light and reflections matter most.

The next chapter turns from the world around the performers to the performers themselves: capturing their movements and faces, altering them, and creating digital humans.

CHAPTER 11

Performances, Faces and Digital Humans

The human face is the most scrutinised image in cinema. Audiences read tiny movements of the eyes and mouth with extraordinary precision, and they notice immediately when something is wrong. For that reason, digital faces and bodies have always been the hardest problem in visual effects, and they remain the most sensitive area of AI filmmaking, technically, ethically and legally. This chapter covers the tools that capture performances and transfer them to digital characters, that change what a performer says or what language they speak, that change a performer's apparent age, and that create digital humans. It covers how these tools work, how productions have used them, and where the lines are.

A short history of performance capture

Capturing a performer's movement to drive a digital character is not new. Motion capture, using markers on a suit tracked by arrays of cameras, became established in film and games in the 1990s and 2000s. Performance capture, which records the face along with the body, enabled celebrated digital characters in the early 2000s and has been refined since. Facial capture used head-mounted cameras and markers or painted dots on the face. These systems produced remarkable results but required specialist stages, suits, cameras and teams, and substantial animation work afterwards to turn the data into a finished character.

Machine learning changed this in two stages. First, markerless capture: systems that estimate body pose, hand pose and facial expression from ordinary video without markers. Second, generative transfer: systems that take a recorded performance and directly generate a video of a

different character performing it, without building and animating a 3D model at all.

Markerless capture to 3D

The first approach keeps a 3D character in the pipeline. A performer is recorded on ordinary cameras. Software estimates their body, hand and facial motion and outputs animation data that drives a rigged 3D character in an animation package or game engine. The character is then lit and rendered conventionally, or rendered roughly and restyled by a video model.

Autodesk's Flow Studio, formerly Wonder Studio, is an example of a product built around this approach. According to Autodesk, it inserts computer-generated characters into live-action footage and captures face, body and hand motion without markers, exporting editable motion capture, camera tracks, masks and clean plates to tools such as Maya, Blender and Unreal Engine. In 2026 it added character adapter training to keep a character's identity consistent across shots.

The advantage of this approach is control. A 3D character can be lit, posed, re-timed and adjusted by animators. The data is editable. The character looks the same from every angle because it is the same model. The disadvantages are the cost of building and rigging a good 3D character, and the long tail of work needed to make it photoreal.

Generative performance transfer

The second approach skips the 3D model. A recorded performance, often from a single ordinary camera, drives a video model that generates a different character performing the same movements and expressions. The character is defined by an image, a reference set or a trained adapter.

Runway's Act-Two, released in July 2025, transferred head, face, body and hand motion from a driving performance video to a character image or video. Kling's motion-control capability, upgraded with its 3.0 model in March 2026, was described by Kling in similar terms, with an emphasis on facial consistency. Other models offered comparable

features, and many leading video models could take a reference performance as one of their inputs by late 2026.

Generative transfer is fast and accessible. An actor in a plain room, a phone on a tripod and an approved character image can produce a performance on screen in minutes. It is particularly strong for stylised and animated characters, where photoreal precision matters less. Its weaknesses are the weaknesses of generation: identity can drift, fine details of the performance can be smoothed or altered, and the output is a video, not editable animation data, so changing it means generating again.

Directing for transfer

Recording a performance for transfer is a form of filmed acting, and the usual directing skills apply. A few practical points matter especially:

- **Match framing.** Record the driving performance in roughly the framing of the final shot. A full-body performance will not transfer well to a close-up character image, or vice versa.
- **Light the face clearly.** Even, frontal light lets the system read expression. Avoid hard shadows across the eyes and mouth.
- **Plain background.** A clear separation between performer and background helps.
- **Match physicality.** The performer's build and movement style should suit the character. A slight performer driving a heavy creature will look wrong unless the movement is adjusted.
- **Avoid occlusion.** Hands in front of the face, hair over the eyes and props crossing the face confuse the system.
- **Perform for the character.** The actor should play the character, not themselves. Their timing and intention come through.

Lip-sync and visual dubbing

A closely related set of tools changes the mouth movements of a performer, real or generated, to match new audio. There are three main uses.

Generated dialogue. For generated characters, lip-sync aligns mouth movement with recorded or synthetic dialogue. Many video models now generate dialogue and lip movement together; others take audio as an input and generate matching video. Dedicated lip-sync tools can also be applied to an existing generated shot.

Dialogue changes in post. Changing a line after shooting, for a story fix, a legal change or a ratings change, traditionally required a reshoot or a careful cutaway. AI tools can now alter the performer's mouth to match a new line, recorded by the same actor. Flawless described its DeepEditor product as changing dialogue or performances in post-production without reshoots.

Visual dubbing. Dubbing a film into another language has always involved a mismatch between the dubbed voice and the original actor's lips. Visual dubbing alters the lip and facial movement to match the new language. Flawless's TrueSync was used to visually dub the feature film *Watch the Skies*, which the company described as the first fully visually dubbed feature, released in US cinemas in 2025 and later on Prime Video. In October 2026 it was reported that Flawless's TrueSync and DeepEditor technology had been sold to a new owner following the company's administration in the UK, a reminder that specialist tools, like models, can change hands or disappear.

Consent and lip-sync

Changing what a real performer appears to say is one of the most sensitive things a production can do. It alters their performance, puts words in their mouth, and may extend to a language they do not speak. Under the SAG-AFTRA agreements described in Chapter 19, digital alteration of a performer's performance generally requires their consent, with a reasonably specific description of the alteration, subject to exceptions for routine post-production changes. Whatever the jurisdiction and agreement, good practice is to obtain written consent for any lip-sync or dialogue change of a real performer, to describe it specifically, to show the result to the performer before it is locked, and to pay for it fairly.

De-ageing and ageing

Making a performer look younger or older has long been possible with prosthetic make-up and, since the late 2000s, with conventional digital techniques, which were expensive and slow and often required extensive frame-by-frame artist work.

Machine learning made it faster and, in some cases, real time. The clearest documented example is Robert Zemeckis's 2024 film *Here*, in which Tom Hanks and Robin Wright played their characters across many decades. According to Metaphysic, the company that provided the AI work, and to accounts published by the trade press and presented at SIGGRAPH 2025, neural networks trained on images of the actors were used to change their apparent age while preserving their live performances. On set, a GPU server received the camera feed, identified each actor and applied a rough age change that was returned to the director's monitor with a delay of a few frames, so the director could judge the performance as it would look in the film. The actors also used a real-time "mirror" to see their younger faces while rehearsing. Final shots were refined afterwards.

The significance of *Here* for a synthetic studio is the method, not just the effect. The performance was real and captured on set; the transformation was applied to it with the actors' participation; and the director could see the result live. That is the pattern this book recommends throughout: real performance, controlled transformation, visible early.

Digital humans

A **digital human** is a photoreal human character that is not a direct recording of a real person in the shot. There are several kinds, with different obligations.

Digital doubles are replicas of a real performer, used for stunts, crowd work, impossible actions or shots where the performer is unavailable. They have been part of visual effects for many years. AI has made them cheaper to create and more convincing.

Digital replicas in the sense of the guild agreements and recent laws are digital representations of a real person's voice or likeness that can be made to perform things the person did not perform. A digital double is one kind. The rules in Chapter 19 apply.

Posthumous performances use a replica of a performer who has died. They raise the sharpest ethical questions and, depending on jurisdiction, legal ones. The 2023 SAG-AFTRA agreement provided that consent given during a performer's life continues after death unless explicitly limited, and that where consent is required and the performer is deceased, it can be given by an authorised representative or, if none can be found, by the union.

Synthetic humans are characters that do not represent any real person: generated faces, digital characters built from scratch. They avoid the consent questions of replicas, but raise the resemblance questions discussed in Chapter 6, and, where they substitute for performers, the labour questions discussed in Chapter 19.

Where the lines are

The technology for face and performance manipulation is the same technology used for deceptive deepfakes. A synthetic studio has to be able to show, at any time, that its use is consented, disclosed and fair. A practical set of rules, which Kestrel Lane follows and which this book recommends as a minimum:

1. **No real person's likeness or voice without written, specific consent.** This includes performers, crew, friends, family and public figures.
2. **No alteration of a real performance without consent** beyond routine post-production adjustments such as colour, cleanup and minor repairs. Lip-sync, dialogue changes, ageing, face replacement and digital doubles all require it.
3. **The performer sees the result** before it is locked, for any alteration of their face or voice.
4. **Data is protected.** Scans, recordings and trained adapters of real people are stored securely, access is limited, and they are deleted when the agreed use ends, unless the performer agrees otherwise.

5. **No deception.** The film's credits and any required disclosures describe the use honestly. Chapter 20 covers disclosure rules, which in some jurisdictions apply to realistic synthetic depictions of people.
6. **No real people in fabricated real events** in work presented as factual, regardless of consent, without clear disclosure.

The decisions on The Salt Road

Kestrel Lane faced three face-and-performance decisions on *The Salt Road*, and their reasoning illustrates the rules above.

The first was whether to generate the keeper's dialogue scenes rather than shoot them, to save a second shoot day. They decided against it. The scenes depended on small, sustained exchanges between two actors, which generation handled poorly, and the actors' performances were the heart of the film.

The second was a line change. After the fine cut, the director wanted to change one of the surveyor's lines to clarify a plot point. The options were to record the new line and cut away during it, or to use lip-sync to alter the actor's mouth to match. The actor was asked, agreed in writing to a single specific change, recorded the new line herself, and approved the result. Had she declined, the cutaway would have been used.

The third was the memory shot of the keeper's wife, a synthetic character, described in Chapter 6. Because the character represented no real person, the questions were resemblance and honesty rather than consent. A resemblance check was run with people unconnected to the project, and the credits note that she is a generated character.

None of these decisions required complex reasoning. Each became simple once the studio had written down its rules in advance.

Building the face pipeline

For productions that rely on face and performance work, it helps to think of it as a pipeline with stages and checks, like any other department.

Capture: record performances and reference with good light and coverage, and record consent at the same time.

Identity: build reference sets or trained adapters for each digital character, with their own bibles and consent records.

Transfer or transformation: apply the chosen method: markerless capture to 3D, generative transfer, lip-sync, de-ageing.

Review: check identity against the bible, check performance against the original, check for artefacts in eyes, teeth, mouth interior and hairline, where face work most often fails, and check at full resolution and speed.

Approval: the director approves the shot; for real performers, the performer sees it.

Record: log the method, tools, inputs, consent reference and approvals.

Faces and performances are where audiences, performers and regulators look first. Care here protects the whole production.

The next part of the book moves into post-production, starting with the edit.

PART IV

Post-Production

CHAPTER 12

Editing the Synthetic Film

Editing is where films are made, and that has never been truer than with generated material. A conventional shoot produces footage that already shares a space, a time, a light and a cast; the editor shapes it. A synthetic production produces shots that were made separately, by different models, at different times, from different random starting points, and that share only what the director deliberately gave them. The editor's job is to make them into one film. This chapter covers editing workflows for generated and hybrid material: organising media, assembling and cutting, handling continuity, using generative tools inside the edit, and conforming for finishing.

The editor's new position

In a conventional production, the editor usually receives footage after the shoot and works with what exists. In a synthetic production, the editor can, in principle, ask for any missing shot to be generated. That changes the role in two ways.

First, the editor becomes part of production. When a cut needs a reaction shot, an insert or a transition that does not exist, the editor can request it, or generate it, the same day. The boundary between production and post-production becomes porous. This is powerful, and dangerous: a film can be endlessly re-shot in the edit, and never finished.

Second, the editor becomes a guardian of continuity in a way that goes beyond conventional editing. Because generated shots share nothing automatically, the editor is often the first person to see that the character's coat has changed colour, that the light has shifted from left to right, or that the sea is calmer than in the previous shot. The continuity strip from Chapter 6 lives in the edit suite.

Kestrel Lane manages this tension with a simple rule. During the assembly and rough cut, the editor may request new shots freely. After the fine cut is locked, any new shot requires the director and producer to agree, and is logged with its reason. It prevents endless re-generation without preventing improvement.

Organising media

Generated productions create a lot of files. Many of them look alike. Good organisation is the difference between an efficient edit and a miserable one.

Name files by shot. Every generated file is named with its shot number, take or variation number, and version, for example *SR_042_v03_t02.mov*. The generation log links that name to the model, inputs and settings. Never rely on the file names a service gives downloads.

Separate selects from attempts. Keep only selected takes and runners-up in the edit project. Archive other attempts outside the project, with their log entries.

Use metadata. Most editing applications allow custom metadata fields. Add fields for method (shot, generated, captured, composited, transformed), model, reference version and consent reference where relevant. These fields let you search, sort and, later, produce the provenance and disclosure records for delivery.

Normalise formats on ingest. Generated clips arrive in many formats, frame rates, resolutions and compressions. Convert them on ingest to the project's editing format and frame rate, keeping the originals. Mixed frame rates in a timeline cause stutters and conform problems.

Watch the frame rate. Many models generate at 24 frames per second; others at 25, 30 or other rates, or offer a choice. Decide the project frame rate before production and generate at it wherever possible. Converting between frame rates, especially from 30 to 24, introduces judder or blending that is very visible in motion.

Assembly

The assembly is the first full version of the film, with every scripted shot in order. For a synthetic production, it often contains a mix of final generated shots, previs, storyboards and placeholder cards for shots not yet made.

That mix is useful. It lets the editor and director see the whole film early, judge pacing and structure, and identify which missing shots matter. Many planned shots turn out to be unnecessary once the assembly exists. Do not wait for every shot before assembling.

Cutting generated material

Editing generated shots raises some specific issues.

Use the middle. As Chapter 7 noted, generated shots are often weakest at their start and end. Trim to the strongest part.

Cut on motion. A cut during movement, such as a turn of the head, a camera move or an object crossing frame, hides small mismatches between shots. Cuts between static frames expose them.

Use cutaways to bridge. When two shots of a character do not quite match, a cutaway to a detail, an object, a landscape or another character between them resets the viewer's eye. Generated cutaways are cheap; plan them.

Watch for invented action. Models sometimes add action that was not asked for: a character who glances at the camera, an object that moves by itself, a background figure who appears. These are easy to miss when focused on the main subject. Review every shot for them.

Respect screen direction. Generated shots have no knowledge of each other's geography. Check that characters face and move consistently across cuts.

Mind the eyelines. In dialogue scenes, characters' eyelines must meet across the cut. Generated reverse shots often have eyelines that do not match. Specify the direction of the look in the prompt, and correct in the frame or with flips where acceptable.

Hold on performance. Where a shot contains a strong performance, real or transferred, cut for the performance first and the image second. Audiences follow faces.

Generative tools inside the edit

Editing applications began incorporating generative tools in 2024 and 2025, and by 2026 several offered them as standard.

Generative extension adds frames to the start or end of a clip. Adobe introduced Generative Extend in Premiere Pro using its Firefly Video Model, initially adding up to two seconds. Extension is useful for giving a shot a little more breath for a transition or for timing to music. It is not a substitute for a longer take, and every extended section should be checked for drift.

Object removal and cleanup. Removing an unwanted object, a boom microphone, a sign or a passer-by, once a visual effects task, can now often be done in the editor with generative fill tools. Check results at full resolution.

Conversational editing. Several video models of 2026 supported changing a generated clip through follow-up instructions, such as "make the light warmer" or "have her look toward the door", rather than regenerating from scratch. Google described multi-turn editing for its Gemini Omni 1.2 Flash model in November 2026. This is useful for small adjustments, but each edit is a new generation with its own risks, and should be logged as such.

Partner models in the editor. Some editing and creative applications offer access to third-party video models inside the application. Adobe, for example, offered partner models alongside its own Firefly models. Note that a vendor's assurances about its own models, such as Adobe's description of Firefly as trained only on licensed content and commercially safe, do not automatically extend to partner models; Adobe's own documentation said that partner models are not covered by its commercial-safety claim and that users must assess each model's terms. Record which model generated each shot regardless of which application you used.

Transcription and search. Automatic transcription of dialogue and searching footage by content, such as "shots with the lighthouse at night", are now common. They help most in hybrid productions with significant shot footage.

Temp sound and temp music

Editors have always cut with temporary sound and music, to judge pacing before the final mix and score exist. Generative tools make temporary voice and music easy to create. Two cautions apply.

First, temporary material has a way of becoming permanent. If a generated temp voice or temp music cue might end up in the finished film, make it with a tool whose terms permit that use, and record it, from the start. Chapter 14 covers this.

Second, the guild and union rules in Chapter 19 may apply to temporary use of synthetic voices of real performers, or to replacing performers' voices in temp. Do not use a cloned voice of a real person in a temp track without their consent.

Continuity review

Before the cut is locked, run a formal continuity review. Watch the film at full resolution, sequence by sequence, with the character bibles, the look document and the continuity strips. Check:

- character identity, costume, props and marks;
- light direction and colour within scenes;
- weather, time of day and environmental state;
- screen direction and eyelines;
- background elements that change between shots;
- text, signs and logos, which models often render wrongly or create unintentionally;
- anything that looks like a real brand, real person or copyrighted character, which must be removed or cleared.

Record issues against shot numbers, and fix them by regeneration, cutaway, reframing, visual effects or acceptance. Some small

inconsistencies do not matter; audiences watching a story do not see what a continuity reviewer sees. Fix the ones that would distract.

Locking and conforming

When the picture is locked, the edit must be prepared for finishing. The **conform** is the process of rebuilding the edited film from the highest-quality version of every shot, at full resolution and quality, in the finishing system.

For generated material, this raises a specific issue: the highest-quality version of a generated shot may not yet exist. Many productions edit with lower-resolution or draft generations and produce final, upscaled or higher-resolution versions only for shots that are used. The conform then involves:

- producing final versions of every used generated shot, matching the edited timing exactly, using the same inputs and settings recorded in the generation log;
- upscaling where needed, as described in Chapter 13;
- checking that the final version matches the draft that was cut, since regeneration at higher quality can change content;
- relinking the edit to the final versions.

This is where a good generation log pays for itself. Without it, matching the final version to the cut can mean starting again.

Export from the editor an edit decision list or interchange file in a format your finishing system reads, along with reference video of the locked cut. Include metadata: shot numbers, methods and models. The finishing team will need it, and so will the provenance record at delivery.

Cutting The Salt Road

The edit of *The Salt Road* ran for about eight weeks, overlapping with generation. A few episodes show the working pattern.

The first assembly ran nineteen minutes against a target of twelve. The storm sequence, which had been the centre of the studio's planning, ran almost four minutes and was the main problem: every generated shot

was beautiful, and the sequence was boring. The editor cut it to ninety seconds by removing almost every shot that did not include the keeper, the surveyor or the light. The cut shots had cost money and time. They were not missed.

The second problem was the surveyor's arrival. In the script, she drives along the coast road in daylight, and the generated road shots did not match the photographed car interior shots: the sky through the car windows was brighter and bluer than the generated landscape. Rather than regenerate the landscape, the editor reordered the scene so that the interior shots came after the car had passed into the shadow of the headland, and the colourist pulled the window highlights down to match. A story problem had been mistaken for a generation problem, and vice versa.

The third episode was a gift from the generation log. The director wanted a moment of stillness after the storm, and nothing in the selects fitted. The editor searched the log for rejected takes of the calm sea and found a runner-up from the first week, rejected because the sea was too blue. Graded to the look, it became the film's final shot.

At fine cut lock, the film was eleven minutes and forty seconds. Of sixty-eight shots in the production draft, fifty-one survived, and nine new shots had been requested during the edit, most of them cutaways.

Versions for different screens

A film is now often delivered in several versions: the main cut, a trailer, short excerpts for social platforms, sometimes a vertical version for phones. Synthetic productions have an advantage here and a trap.

The advantage is that a vertical or square version need not be a crop of the horizontal one. Key shots can be regenerated or extended at a different aspect ratio from the same key frames and references, or recomposed from the captured and composited elements, so the vertical version is composed for its screen rather than cut down from another. The previs and the asset records make this feasible.

The trap is that every regenerated shot in an alternative version is a new piece of material that must match the original in identity, light and

content, be logged, and be covered by the same consents and disclosures. A vertical version in which the main character looks slightly different, or in which a regenerated background now contains a logo, undoes the care taken over the main cut. Treat alternative versions as productions in their own right, with their own shot inventory, review and QC.

The editor's record

At picture lock, Kestrel Lane's editor produces a **shot inventory**: a list of every shot in the locked cut with its number, duration, method, source files, model or camera, reference versions, consent references and any notes. It is generated mostly from the edit metadata. The inventory feeds four later processes: the conform, the VFX and finishing schedule, the rights and clearance review, and the provenance and disclosure records for delivery. It is the single most useful document a synthetic production creates in post.

The next chapter covers the visual effects work that happens between lock and finishing.

CHAPTER 13

AI VFX: Roto, Relighting, Cleanup and Upscaling

Visual effects work has always been a mix of artistry and drudgery. The artistry is in designing an effect that serves the story and making it believable. The drudgery is in the thousands of hours of rotoscoping, tracking, cleanup and frame-by-frame correction that the artistry depends on. Machine learning has made its largest practical impact on the drudgery, and that impact matters as much to conventional productions as to synthetic ones. This chapter covers the AI tools that changed visual effects work between 2023 and 2026: segmentation and rotoscoping, relighting from estimated material properties, cleanup and paint, character insertion, upscaling and frame interpolation, and how to bring them into a compositing workflow that a finishing team can trust.

The compositing foundation

Before discussing tools, it helps to recall what compositing involves. A composite combines elements from different sources into a single image. To do that, the compositor needs:

- **mattes**, which define which pixels belong to each element;
- **tracking** data, which describes how the camera and objects move, so elements stay locked together;
- **colour and light matching**, so elements look as if they share the same light and camera;
- **edge and depth integration**, so elements overlap correctly, with appropriate softness, blur and atmosphere;
- **grain and lens matching**, so all elements share the same texture and optical character.

Every AI tool in this chapter helps with one or more of these. None of them removes the need for the compositor's judgement about whether the result is believable.

Segmentation and rotoscoping

Rotoscoping, which means drawing mattes by hand around elements frame by frame, has long been one of the most labour-intensive tasks in visual effects. AI segmentation tools that find and track objects through video have made much of it faster.

Meta's Segment Anything Model family is the best-known example. SAM 3, released on 19 November 2025, could detect, segment and track objects in images and video using text prompts, such as "the woman in the red coat", or visual prompts, such as a click or a box, according to Meta. Meta published code for it, and third-party video tools soon began offering it for rotoscoping. Many other segmentation and matting tools exist, both open and commercial, and editing applications include their own.

AI segmentation produces good mattes quickly for well-defined subjects. It still struggles with fine, semi-transparent detail, such as hair, motion blur, smoke and glass, which is exactly where mattes matter most for a convincing composite. A practical workflow:

1. Generate a matte with an AI tool.
2. Review it at full resolution, especially at edges and in motion.
3. Refine edges with matting tools designed for hair and transparency.
4. Correct failures by hand in the frames where they occur.
5. Keep the matte as a separate file, versioned with the shot.

Segmentation also has a less obvious use in synthetic productions: isolating elements of a generated shot so they can be adjusted separately. A generated shot with a character whose coat has the wrong colour can be fixed by segmenting the coat and grading it, rather than regenerating the whole shot.

Depth, normals and material estimation

Many AI tools estimate from ordinary images properties that once required special capture: depth (how far each pixel is from the camera), surface normals (which way each surface faces), and material properties such as base colour, roughness and specularity. These estimates are useful throughout compositing.

Depth enables atmospheric effects, such as fog, haze and depth of field, that increase with distance; correct layering of elements; and depth-conditioned generation, as Chapter 8 described.

Normals and materials enable relighting.

Relighting

Changing the lighting of a shot after it has been photographed or generated was traditionally very difficult. It required a 3D model of the scene or extensive painting. AI relighting estimates the geometry and materials of what is in frame and uses them to apply new light.

Beeble's SwitchLight is an example designed for production. According to Beeble, SwitchLight 3.0 converts video into temporally consistent physically based rendering passes, which are normals, base colour, roughness, specular and metallic, running locally in its desktop application at up to 4K, with 16-bit EXR output. Those passes can be brought into a compositing package or, through a plugin, into Unreal Engine, where the footage can be relit interactively with the engine's lighting.

Video-to-video models offer a different route to relighting: describing the new light in a prompt and letting the model transform the shot. Runway listed relighting among Aleph's capabilities in 2025. This is faster and requires less setup, but gives less precise control and may alter other details of the shot.

Relighting is particularly valuable in hybrid productions, where a performer shot under one light must match a generated or captured environment with another. It is also where consent questions can arise: relighting changes the appearance of a real performer, and while modest relighting is a routine part of post-production, dramatic changes may be

the kind of alteration a performer should know about. Kestrel Lane's keeper was told in advance that three of his shots would be relit to match generated lightning, and saw the results.

For *The Salt Road*, the relighting workflow for the lightning shots was:

1. Shoot the keeper at the window with soft, neutral light and a dark background.
2. Generate PBR passes from the plate with a relighting tool.
3. In compositing, animate a hard, blue-white light from outside the window, synchronised to the generated lightning flashes.
4. Combine with the original plate, keeping skin texture from the original and taking light from the relit pass.
5. Review with the director and the actor.

Cleanup and paint

Removing unwanted elements from a shot, such as rigs, wires, crew reflections, modern objects in a period scene, logos and signs, is a staple of visual effects. AI-powered inpainting and video object removal now handle many simple cleanup tasks well, and many editors, compositing tools and video models include them.

The challenge in video is temporal consistency: the filled area must look right in every frame and must not flicker or swim as the camera moves. Results vary widely with the shot. Static cameras and simple backgrounds work well. Complex moving backgrounds, large removals and areas the camera reveals over time are harder.

Cleanup in generated material has a specific use: removing the things models invent. Unwanted text and pseudo-text on signs and screens, logos that resemble real brands, extra figures in backgrounds, malformed objects at the edge of frame. A cleanup pass on every generated shot that reaches the final cut is good practice.

Character insertion and replacement

Inserting a computer-generated character into live-action footage, or replacing a performer with a digital character, combines several of the

tasks above: tracking the camera, capturing the performer's motion, removing the performer from the plate, rendering the character, and compositing it in matching light.

Autodesk's Flow Studio, described in Chapter 11, was built to automate much of this chain: from a single shot it could track the camera, capture the performance, generate a clean plate with the performer removed, and output render passes and scene files for further work in 3D tools. Video models with reference and motion transfer offer a generative alternative, producing the replacement directly as video.

For hero characters in finished work, many productions combine the approaches: AI tools for the tracking, capture, clean plate and first pass, with artists refining the result in conventional software.

Generative effects elements

Effects elements, such as fire, smoke, water splashes, explosions, debris, rain, snow and dust, have traditionally been simulated in 3D or filmed as practical elements against black. Video models can generate them, either as complete shots or as isolated elements to be composited.

Netflix's account of its use of generative AI in *El Eternauta* in 2025, a building collapse sequence produced with its visual effects studio, is the best-known disclosed example of generated effects in a major series. According to Netflix's co-chief executive, the sequence was completed considerably faster than with traditional visual effects workflows and at a cost that otherwise would not have fit the show's budget. Netflix did not name the specific tools.

When generating effects elements for compositing:

- generate against a plain black or neutral background, so the element can be extracted;
- match the camera angle, lens and motion of the plate;
- match the light direction and colour;
- generate at sufficient resolution and bit depth for compositing, or upscale carefully;
- check for physical plausibility at the scale of the shot.

Upscaling

Many video models generate at resolutions below the final delivery resolution, offering upscaling as a separate step. Google, for example, described drafting at 360p or 720p and upscaling selected clips to 1080p or 4K for its Gemini Omni 1.2 Flash model. Dedicated upscaling tools also exist, both commercial and open.

AI upscaling does not recover detail that was never captured. It generates plausible detail consistent with the lower-resolution image. That is useful, and it is also a source of problems:

- **Invented detail** may not match between shots, between frames, or with the original intention: skin texture, fabric weave and foliage can be rendered differently each time.
- **Over-sharpening** gives an artificial, crisp look that clashes with photographed material.
- **Face alteration**: some upscalers subtly change facial features, which can break character consistency.
- **Temporal instability**: detail that flickers or swims between frames.

Good practice is to upscale only the final selected and conformed shots, to use the same upscaler and settings for all shots in a sequence, to compare faces before and after against the character bible, and to add grain consistently afterwards so that upscaled and photographed material share a texture.

Frame interpolation and retiming

AI frame interpolation generates in-between frames to change frame rate or create slow motion. It is useful for smoothing motion in generated clips, for converting between frame rates, and for slow-motion effects from footage shot at normal speed. Its typical failures are warping and smearing around fast motion, occlusions and thin objects. Use it carefully, check results at full speed, and prefer generating at the project frame rate from the start.

Integration and grain

The last step in any composite involving generated material is integration: making every element share the same optical character. This means matching black levels, contrast and colour; adding lens effects such as slight chromatic aberration, vignetting and halation consistently; matching depth of field and motion blur; and adding a single, consistent grain structure across the whole frame.

Grain matters more than most beginners expect. Generated images are typically clean and smooth; photographed images carry sensor noise or film grain. A composite in which one element is clean and another grainy looks wrong even when everything else is right. Many productions remove or reduce noise from all elements, then add a single, consistent grain to the finished composite. Chapter 15 returns to grain as part of the finishing recipe.

Planning and bidding VFX work

AI tools have changed the economics of individual VFX tasks, but not the need to plan them. For each VFX shot, the production should still record what the shot needs (clean plate, roto, relight, insertion, effects elements, comp), what tools and methods will be used, how long it will take, and who will approve it.

When working with outside VFX vendors, be explicit about AI use in both directions. Tell the vendor which elements of your material were generated and with what tools, so they can plan for its characteristics, such as bit depth and compression. Ask the vendor which AI tools they will use on your material, where those tools run, and what their data terms are, so that your provenance record and any partner requirements are met. Some studios' guidance for production partners, as Chapter 17 describes, requires exactly this.

Estimates for AI-assisted VFX work are harder than for conventional work, because AI tools can be very fast when they succeed and slow when they fail and the shot falls back to manual methods. Bid with that variance in mind: estimate each shot's AI path and its manual fallback, and agree in advance how overruns will be handled.

Quality control for AI VFX

Because AI tools can fail in subtle ways, quality control needs to be systematic. For every VFX shot, review:

- **at full resolution**, on a calibrated monitor;
- **at full speed and frame by frame**, since some failures are only visible one way or the other;
- **edges**: mattes, hair, motion blur, transparency;
- **faces**: identity, eyes, teeth, mouth interior, hairline;
- **temporal stability**: flicker, swimming texture, popping detail;
- **invented content**: text, logos, extra figures, objects;
- **light and colour** relative to neighbouring shots.

Record the review result and the reviewer in the shot's record. In larger productions, the VFX supervisor signs off each shot.

Data handling

Many AI VFX tools run in the cloud. Uploading unreleased footage, especially footage of performers, to a third-party service raises confidentiality and data protection questions. Studio guidance for production partners, such as Netflix's guidance described in Chapter 17, has specifically asked partners to use tools that do not retain or train on production data and to work in secured environments. Prefer tools that run locally, such as those that advertise on-device processing, for sensitive material, or use enterprise agreements with clear data terms. Record which tools processed which footage.

The next chapter turns from images to sound: voice, dialogue and music.

CHAPTER 14

Voice, Dialogue and Music With Consent

Sound is half of a film, and in synthetic productions it is often the half that is neglected. That is a mistake in two directions. Creatively, sound carries emotion, continuity and meaning that pictures alone cannot; a mediocre generated image with excellent sound will hold an audience better than the reverse. Legally and ethically, voice and music are where some of the sharpest questions about consent, licensing and labour arise. This chapter covers voice and dialogue, synthetic and cloned voices, licensed voice marketplaces, generated music and its licences, and how to build a soundtrack for a synthetic film. It deliberately stops short of detailed audio engineering, which is the subject of the third AWC Field Manual, *Signal Rooms*.

The soundtrack in layers

A film soundtrack is traditionally built from four families of sound:

- **dialogue**, including on-camera speech, off-camera lines and voice-over;
- **effects**, including hard effects synchronised to picture, Foley such as footsteps and cloth movement, and backgrounds or ambiences;
- **music**, including score and source music;
- **silence**, which is used deliberately.

Each family is edited separately and combined in the mix. A synthetic production should keep that structure even when some layers are generated, because it allows each layer to be controlled, replaced and cleared independently.

Generated sound with picture

As Chapter 3 described, several leading video models of 2025 and 2026 generate synchronised sound with the picture: ambience, effects, music and dialogue. This is very useful. A generated shot with appropriate sound is easier to judge, and for short-form work the generated sound can sometimes be used as is.

For longer or more serious work, treat generated sound with picture as a draft:

- it is usually mixed into a single track, so dialogue, effects and music cannot be separated cleanly;
- it varies from shot to shot, so ambiences do not match across cuts;
- its dialogue performance is hard to control;
- its rights status follows the video model's terms, which may or may not permit separate use of the audio.

Most productions therefore replace or rebuild the sound in post, using generated sound only as reference or as individual elements where it works. Tools that separate mixed audio into stems, such as dialogue, music and effects, can help extract useful parts; their results vary, and *Signal Rooms* covers them in detail.

Dialogue: recorded, synthetic or cloned

There are three ways to get dialogue into a synthetic film.

Recorded performance. A human actor performs the lines, recorded in a quiet room or studio. For any character whose voice matters, this remains the strongest choice. Actors bring interpretation, timing and life. The recording can drive the picture through lip-sync or performance transfer, as Chapter 11 described. The actor is engaged under an ordinary performer agreement, with the appropriate union terms where they apply.

Synthetic voice. A text-to-speech model generates speech from text in a voice provided by the service, one that does not belong to a specific real person. Modern synthetic voices are expressive and can follow direction about emotion, pace and emphasis. They are suitable for many uses:

narration, minor characters, temp tracks, prototypes and accessibility versions. The voice's licence comes from the service's terms; check that they permit your use, including commercial distribution.

Cloned voice. A model is trained or prompted with recordings of a specific person's voice to generate new speech in that voice. Cloning can be used to extend a performer's work, for example to fix a line after a shoot without bringing them back to a studio, to dub them into another language, or to create a character voiced by a performer who cannot record new material. It is also the basis of voice fraud and impersonation.

Consent for voices

The rule for cloned voices is simple and absolute: **never clone a real person's voice without their explicit, written, specific consent for the intended use.** This applies to performers, crew, friends, family and public figures. It applies to temp tracks as well as finished work.

Specific consent means the agreement names the project, describes the use (for example, "new lines for the character Mara in the film *The Salt Road*, up to twenty lines, for dialogue fixes"), states the media and territories, states the duration, sets payment, gives the performer approval over the result, and says what happens to the voice model afterwards.

The union agreements and laws discussed in Chapter 19 treat voice replicas alongside likeness. The SAG-AFTRA agreements cover digital replicas of a performer's voice. Tennessee's ELVIS Act, in force since July 2024, specifically protects voice. Other jurisdictions protect voice through publicity, personality or passing-off law. Even where no specific rule applies, the ethical position is clear.

Licensed voice marketplaces

A newer option sits between synthetic and cloned voices: marketplaces where well-known people, or the estates of people who have died, license their voices for approved uses. ElevenLabs launched an "Iconic Marketplace" of this kind. According to the company, each commercial

use must be approved by the talent or an authorised rights holder, with terms, territory, duration and commercial rights set in a licensing agreement. Reports in 2026 noted that several living actors had licensed their voices through it, and raised the question of what consent means for historical figures who have died, where only their estates can agree.

Marketplaces like this offer a consented route to recognisable voices. Before using one, check that the approval process covers your specific project, that the licence terms fit your distribution, and that you are comfortable with how the estate consent for deceased figures works. Disclosure obligations may still apply.

Dubbing and localisation

AI tools have changed dubbing more than almost any other area of sound. Translation, synthetic voices in many languages, voice cloning that keeps an actor's own voice in another language, and visual dubbing that changes lip movement to match, as Chapter 11 described, can be combined into an end-to-end localisation pipeline.

These tools can make international distribution affordable for small productions, and they raise serious labour questions, since dubbing has long been skilled human work, with its own performers, directors and translators. Productions should consider carefully which parts of localisation to automate, obtain consent from the original performers for any use of their voices or faces in other languages, and follow union rules where they apply. Quality varies by language; native speakers should review every localised version.

Music

Music for a synthetic film comes from the same sources as music for any film, plus one more.

Commissioned score. A composer writes original music for the film. This remains the best option for most narrative work, because a score written to picture serves the story in ways that library music and generated music rarely match. The composer's agreement sets ownership and licensing.

Licensed music. Existing recordings are licensed from their rights holders, usually requiring both a synchronisation licence for the composition and a master use licence for the recording. Production music libraries offer pre-cleared music under simpler licences.

Generated music. A music model generates music from text descriptions, reference audio or other inputs.

Music generation went through a turbulent period. Major record companies sued leading music-generation companies in 2024. In late 2025, several of those disputes turned into settlements and licensing agreements, according to reports by the Associated Press and Reuters and announcements by the companies. Universal Music Group settled with Udio in October 2025 and announced a new licensed platform; Warner Music Group announced agreements with Udio and Suno in November 2025, including licensed models to be launched in 2026. Reports described some resulting services as more restricted than before, for example limiting downloads. Not every dispute had settled by early 2026. Separately, ElevenLabs launched Eleven Music in August 2025, describing it as cleared for broad commercial use subject to plan terms, and announced agreements with the independent label network Merlin and the publisher Kobalt.

The practical lessons for a studio are:

- **The terms are the licence.** Whether you can use generated music in a film distributed on streaming services, in cinemas or on television depends on the specific service, plan and date. Terms have changed, and some plans have prohibited particular uses such as streaming distribution.
- **Keep the evidence.** Record the service, plan, date, prompt and output for every generated cue, and keep a dated copy of the terms.
- **Watch for resemblance.** A generated cue that sounds like a specific existing recording or composition is a problem regardless of the service's terms. Do not prompt for named artists' styles.
- **Copyright may be thin.** As Chapter 18 explains, purely generated music may not be protected by copyright in some jurisdictions. That

matters if you want to license the soundtrack or prevent others from using it.

- **Consider a hybrid.** Many productions use generated music as a sketch or temp score and have a composer write the final score, or have a composer shape, arrange and perform over generated material, which strengthens both quality and human authorship.

Effects and ambience

Effects and ambience are the lowest-risk area for generated sound. Footsteps, wind, rain, doors, machinery and crowds rarely raise consent or resemblance issues, and generated effects can be very good. Recorded sound remains valuable: a real recording of the actual location has a specificity that generated ambience lacks. Many sound designers combine libraries, field recordings and generated elements.

For *The Salt Road*, Kestrel Lane recorded wind, sea and the creaking of the real lighthouse during the capture session, which became the foundation of the soundtrack. Generated effects filled specific needs: the roar of the storm at its height, the crack of lightning, the hum of the great lens's motor.

Directing synthetic voices

When a synthetic voice is used, for narration, a minor character or a temp track, it still needs direction. Text-to-speech models of 2025 and 2026 accepted a range of direction, through text instructions, tags or reference recordings, and responded to it unevenly. A few practices help.

Write for the ear. Scripts written for reading often sound stiff when spoken. Read every line aloud before generating it, and rewrite for natural speech: shorter sentences, contractions, clear stress.

Direct with context. Describe the situation and the intention rather than the emotion alone: "speaking quietly so as not to wake someone" works better than "sad".

Control pace with punctuation and breaks. Commas, full stops and explicit pauses where the tool supports them shape rhythm more reliably than instructions.

Generate several readings. As with picture, generate a small batch and choose. Splice the best phrases from different readings if necessary, as editors do with recorded dialogue.

Match the room. Synthetic voices are usually dry and close. In the mix, give them the acoustic space of the scene, or they will sound pasted on.

Keep the record. Log the voice, the model, the text, the settings and the selected reading, as for any generated asset.

Where the synthetic voice stands in for a performer who will later record the line, mark it clearly as temp in the edit, so it is not forgotten.

The mix

All layers come together in the mix. The principles are those of any film mix: dialogue clear and intelligible, effects supporting the picture, music supporting the emotion without fighting the dialogue, levels consistent across scenes, and the whole mix meeting the loudness targets for its destination, which Chapter 15 lists.

Two issues are specific to synthetic productions. First, generated audio elements often have different tonal character, noise floors and spatial qualities from recorded ones, and need more processing to sit together. Second, generated audio may contain artefacts, such as metallic tones, smeared transients and odd reverberation, that are not obvious on laptop speakers but are on good monitors. Mix and review on accurate speakers.

Accessibility

AI tools have made two important accessibility deliverables far easier to produce: captions and audio description.

Automatic transcription produces a first draft of captions and subtitles quickly and, for clear dialogue in common languages, accurately. It still makes errors with names, accents, overlapping speech, technical terms and invented words, and it does not by itself produce good captions, which also need correct timing, line breaks, speaker identification and descriptions of significant sounds. Have every caption file reviewed and corrected by a person, ideally someone experienced in captioning.

Audio description narrates the visual content of a film for blind and partially sighted audiences, in the gaps between dialogue. Language models can draft descriptions from scripts and shot lists, and synthetic voices can perform them. Description is a craft, with established guidelines about what to describe, when, and in what words; a draft from a model is a starting point for a describer, not a finished track. Where blind and partially sighted viewers can review the result, their feedback is the best test.

Many platforms and broadcasters require captions, and some require audio description, as part of delivery. Plan both into the schedule; they are easier to do well when the film's records, such as the script, shot inventory and character bibles, are available to the people making them.

A handoff to Signal Rooms

This chapter has covered sound at the level a director and producer need. The engineering behind it, which includes stem separation, voice models and their training, AI-assisted mixing and mastering, spatial audio, loudness measurement and delivery, and the practical setup of a sound room that uses these tools well, is the subject of the third AWC Field Manual, *Signal Rooms*.

The next chapter turns to the final stage of picture: colour, finishing and delivery.

CHAPTER 15

Colour, Finishing and Delivery

A film is not finished when the edit is locked. It is finished when a file that meets the destination's technical specification has been delivered and accepted. Between those two moments lie colour grading, finishing, the final mix, quality control and the creation of deliverables, a stage that independent filmmakers often underestimate and that synthetic productions find particularly demanding, because generated footage arrives with technical characteristics quite different from camera footage. This chapter covers colour management, grading generated and hybrid material, bit depth and its consequences, high dynamic range, delivery formats and specifications, loudness, and quality control.

Why generated footage is technically different

Camera footage from a modern cinema camera is typically recorded in a log or raw format, with high bit depth, wide dynamic range and wide colour gamut. It looks flat and desaturated straight from the camera, because it is preserving information for the colourist to shape. It is scene-referred: its values represent light in the scene rather than an image ready for a display.

Generated footage is usually the opposite. Most video models deliver clips that are:

- **display-referred**, already graded for a standard display, with contrast and colour baked in;
- **standard dynamic range**, usually Rec. 709 or sRGB;
- **8-bit** in most cases, meaning 256 levels per channel, compared with 10, 12 or more for camera footage;
- **heavily compressed** with consumer codecs such as H.264 or H.265;

- **sometimes lower resolution** than the delivery format, requiring upscaling.

These characteristics limit what a colourist can do. Pushing an 8-bit, compressed, display-referred image hard in a grade produces banding in skies and gradients, blocky artefacts in shadows, and colour breaks. The image has less room to move.

Some services and open-weight workflows offered higher-quality outputs by 2026, including higher bit depth and, in at least one announced case, HDR output: Kling described 4K HDR output for Kling 4.0 in October 2026, and ShengShu described 10-bit 4K output for a preview of Vidu Q4, according to a third-party comparison. Check what your chosen models offer, and request the highest quality available for final shots.

A colour-managed pipeline

The answer to mixed sources is a colour-managed pipeline: a defined working colour space into which every source is converted on input, in which all grading happens, and from which every output is produced through defined transforms.

The most widely used system in film and television is ACES, the Academy Color Encoding System, maintained by the Academy of Motion Picture Arts and Sciences and its partners. ACES 2.0 was released on 4 April 2025, with redesigned output transforms that convert scene-referred working data to images for specific displays, combining tone scale, chroma compression and gamut compression, according to its documentation. Many grading and compositing applications support ACES, as do some real-time engines.

In an ACES pipeline:

- camera footage enters through an input transform specific to the camera;
- generated footage, which is display-referred, enters through an inverse of an appropriate display transform, which converts it into scene-referred working space as best it can;
- grading and compositing happen in the working space;

- outputs for each display, such as SDR Rec. 709, HDR PQ or cinema P3, are produced through the matching output transforms.

Converting display-referred generated footage into a scene-referred space is not magic: it cannot recover highlight or shadow detail that the model did not produce. But it puts all footage on a common footing, so that a grade affects camera and generated footage in the same way, which is the main requirement for making them match.

Whatever system you use, the essential practice is to decide the pipeline before production, test it with real samples of every source, and apply it consistently. Kestrel Lane tested their pipeline in the first week of *The Salt Road*, with a camera test from the lighthouse and generated test shots from each of the two video models they planned to use.

Grading generated and hybrid material

Grading a synthetic film has three goals, in order.

Match. Bring every shot in a scene to the same exposure, contrast, colour balance and saturation, so the scene feels continuous. Generated shots from different sessions and models differ more than shots from one camera on one day. Matching usually takes most of the grading time.

Look. Apply the film's look across all shots: the palette, contrast, texture and mood defined in the look document. A look applied across everything is a powerful unifier.

Emphasis. Guide the eye within each shot, with selective adjustments, vignettes and windows, as in any grade.

Practical tips for generated material:

- **Grade gently.** Generated footage has less room to move. Large adjustments cause banding and artefacts. If a shot needs a big change, consider regenerating it closer to the target.
- **Fix colour with masks.** Use segmentation, as Chapter 13 described, to isolate elements that need different treatment, such as a coat that has drifted in colour.
- **Watch skin tones.** Generated skin can be unnaturally smooth and uniform. Grading can add some variation, and grain helps.

- **Denoise and degrain first.** Remove compression noise and any inconsistent grain before grading, then add a single grain at the end.
- **Use reference monitoring.** Grade on a calibrated display in a controlled environment. Generated footage often looks acceptable on a laptop and shows its problems on a good monitor.

Bit depth, banding and dither

Banding, visible steps in smooth gradients such as skies, fog and out-of-focus backgrounds, is the most common technical problem in finished synthetic films. It comes from low bit depth, compression and grading.

To reduce it:

- request the highest bit depth available for final generations;
- upscale and process in high bit depth, with all intermediate files at 10 bits or more, ideally in floating-point formats such as OpenEXR for visual effects;
- grade gently on problem shots;
- add grain or dither before final encoding, which breaks up bands;
- check gradients on a good display, at full resolution, after the final encode, since encoding itself can introduce banding.

High dynamic range

High dynamic range delivery, in formats such as Dolby Vision and HDR10 using the PQ transfer function, is common for streaming platforms. HDR can make images more vivid and realistic, but it exposes the limitations of source material more harshly.

For synthetic productions, HDR raises a question: since most generated footage is SDR, how can it be delivered in HDR? Options include:

- generating with models that offer HDR or higher bit depth output;
- converting SDR to HDR with inverse tone mapping, carefully, accepting that highlights will not have the detail of true HDR;
- delivering the film in SDR only, if the destination allows.

Netflix's published guidance for Dolby Vision mastering, for example, specifies P3-D65 mastering in PQ at 12-bit for UHD Dolby Vision deliverables with per-frame metadata. Meeting such specifications with converted SDR material is possible, but the result must be checked carefully, and productions should confirm with the platform how it treats upconverted material.

Resolution and upscaling in finishing

If generated shots are below delivery resolution, they must be upscaled, as Chapter 13 discussed. In finishing, the key is consistency: all upscaled shots should be processed with the same tool and settings, compared with camera footage at delivery resolution, and integrated with the same grain. Platforms may have rules about how much upscaled material is acceptable in certain delivery tiers; check before committing.

Grain and texture as finishing

Grain was discussed in Chapter 13 as a compositing tool. In finishing it becomes a creative and technical unifier. A consistent grain structure across the whole film:

- makes generated, captured and photographed material feel like one medium;
- hides small differences in detail and smoothness between sources;
- breaks up banding;
- gives the image texture that many audiences associate with cinema.

Choose grain with intention. Fine grain at a low level unifies without drawing attention. Heavy grain makes a stylistic statement. Test it at the final delivery resolution and after the final encode, because compression can smear grain into noise or remove it entirely.

Delivery formats and specifications

Every destination has its own specification. Common categories:

Online platforms generally accept a high-quality file in a standard codec, such as H.264 or H.265 at a high bit rate, or a mezzanine codec, and re-encode it.

Festivals generally request a screening file, often H.264 at high quality, and for selected films a Digital Cinema Package (DCP), which is the standard format for cinema projection.

Streaming services and broadcasters have detailed specifications. Netflix's published delivery requirements for its branded productions, for example, call for an Interoperable Master Format package conforming to SMPTE ST 2067-21 Application #2E, with specific requirements for video, audio, subtitles and metadata. Broadcasters have their own specifications, often including technical checks for colour gamut and loudness.

Archive masters should be kept at the highest quality: an uncompressed or lightly compressed master in the working colour space or the highest display version, plus separate audio stems, textless versions and project files.

Read the specification before finishing begins, not at the end. Some requirements, such as frame rate, resolution, aspect ratio, safe areas for titles and audio channel configuration, affect decisions made much earlier.

Festival deliverables

Festivals commonly ask first for an online screener and, if a film is selected, for an exhibition copy, often a Digital Cinema Package for cinema projection, sometimes alongside other formats. A DCP has its own colour space, frame rate options and audio configurations, and it is usually made by a specialist or with dedicated software from a high-quality master. If festivals are part of the plan, find out the requirements of the most likely festivals early, budget for a DCP, and test it in a cinema if you can. Generated material that looked acceptable on a monitor can reveal banding, noise and softness on a large screen, and it is better to find this before a premiere than during one.

Loudness

Loudness specifications ensure consistent volume between programmes. They are measured with standard methods defined in ITU-R BS.1770, and targets vary by destination:

- the European Broadcasting Union's recommendation R 128 sets a programme loudness target of minus 23 LUFS for broadcast;
- Netflix's published specification for its branded productions sets dialogue-gated loudness at minus 27 LKFS, plus or minus 2 LU, with true peak not exceeding minus 2 dBTP;
- online platforms often normalise playback to their own targets, which differ by platform.

Mix to the destination's specification, measure with a compliant meter, and keep a loudness report with the deliverables. *Signal Rooms* covers loudness in depth.

Quality control

Quality control, or QC, is the systematic check of a finished deliverable against its specification and for technical and content faults. Platforms often run their own QC and reject deliverables that fail. A production should run its own first.

A QC pass for a synthetic film should check:

- **technical compliance**: format, resolution, frame rate, colour space, bit depth, audio configuration, loudness;
- **picture faults**: banding, compression artefacts, flicker, dropped or duplicated frames, dead pixels, edge artefacts;
- **generated-content faults**: identity drift, invented text, logos, extra limbs or figures, physics errors not caught in the edit;
- **audio faults**: clicks, distortion, sync drift, missing channels, artefacts in generated audio;
- **content and compliance**: credits, disclosures, legal lines, subtitles and captions;

- **metadata and provenance**: that required metadata, content credentials and disclosure records are present and correct, as Chapter 20 describes.

Record the QC results, fix the faults, and re-check.

Archiving

A finished film should be archived so that it can be re-versioned, restored and, if necessary, defended years later. For a synthetic film, the archive should include more than the masters:

- the picture masters in the working space and in each delivered version;
- audio stems and the final mix sessions;
- the locked edit and the conform data;
- the shot inventory, generation logs and asset records;
- the reference sets, character bibles, look documents and adapters, with their base model versions;
- captured frames and trained captures;
- the template graphs and their versions;
- local copies of open-weight models used, where licences allow;
- dated copies of service terms;
- consent records, permissions and the provenance report.

Store the archive in at least two places, one offline or off-site, and check it periodically. Generated material cannot always be regenerated: models are withdrawn, as Sora was, and services change. The archive may be the only copy of the film's raw material that will ever exist.

Kestrel Lane's finishing recipe

For *The Salt Road*, Kestrel Lane wrote a one-page finishing recipe before production began and updated it as they learned. Its final version read, in summary:

- project frame rate 24 fps, delivery 4K UHD SDR and an HD festival screener; HDR not attempted;

- ACES working space; camera input transforms for photographed material; inverse Rec. 709 output transform for generated material;
- final generations requested at the highest quality available; all intermediate files 16-bit;
- one upscaler and one set of settings for all generated shots, applied after conform;
- denoise all generated material lightly; add a single fine grain across the film at 4K before encoding;
- dialogue-gated loudness to the strictest target among planned destinations, with versions adjusted for others;
- QC at 4K on a calibrated display and on a television, with the shot inventory to hand.

A recipe like this is a small document that saves a great deal of confusion. Write yours early.

The next part of the book steps back from individual stages to the studio as a whole: how to connect these stages into a pipeline, and how to cost and staff it.

PART V

The Studio Pipeline

CHAPTER 16

Node Graphs and Agentic Pipelines

A first film can be made by clicking through web interfaces. A studio cannot run that way for long. When a project has hundreds of shots, several characters with trained adapters, multiple models, a dozen reference sets and a delivery specification, the work has to become repeatable: the same inputs should produce the same kind of output, every step should be recorded, and routine work should not depend on someone remembering which buttons they pressed last month. This chapter covers the two technologies that turned generative filmmaking from a craft of individual sessions into a pipeline: node-based workflow graphs, and AI agents that operate those graphs and other tools on a person's behalf. It ends with the asset, version and review systems that hold a pipeline together.

Why pipelines

Visual effects and animation studios have run on pipelines for decades. A pipeline defines how work moves between departments: what file formats, naming conventions, folder structures, software versions and handoff checks apply at each stage. Its purpose is not bureaucracy. It is to make sure that when a shot moves from layout to animation to lighting to compositing, nothing is lost, everyone works on the right version, and problems can be traced.

A synthetic studio needs the same, plus three things particular to generation:

- **reproducibility**: the ability to regenerate a shot from its recorded inputs, models and settings;
- **substitutability**: the ability to swap a model for another without rebuilding everything around it;

- **provenance**: a reliable record of what produced each asset, for rights, disclosure and repair.

Node graphs and agents, used well, provide all three.

Node-based workflows

A node-based workflow represents a process as a graph. Each node does one thing: load an image, load a model, encode a prompt, apply a reference, sample, decode, upscale, save. Connections between nodes carry data from one step to the next. The graph as a whole describes the entire process, from inputs to outputs, and can be saved, shared, versioned and run again.

Compositing software has used node graphs for decades, so the idea is familiar to visual effects artists. In generative AI, the open-source application ComfyUI became the best-known node-based environment, initially for image generation with open-weight models and later for video, audio and 3D. Its graphs can include open-weight models running on local hardware and, through so-called partner nodes, commercial models accessed through their providers' services. For example, ComfyUI added nodes for Alibaba's Wan 3.0 in August 2026, supporting text-to-video, image-to-video and reference-to-video with synchronised audio, according to its documentation. Commercial platforms also offer their own node or workflow builders; Runway, for example, offered saved Workflows in its product.

What graphs make possible

Combining controls. A graph can combine a start frame, references, a trained adapter, a depth pass from previs, a pose sequence from a recorded performance and a camera path in one generation, as Chapter 8 described. Many such combinations are difficult or impossible in simpler interfaces.

Chaining stages. A single graph can generate a key frame with an image model, animate it with a video model, upscale the result, apply a segmentation to isolate the character, grade it to a reference and save it with a standard name. What would be six manual steps becomes one run.

Batching. The same graph can run across a list of shots, each with its own inputs, overnight.

Reproducibility. Because the graph records every node, model and setting, and because many tools embed the graph in the output file's metadata, a shot can be regenerated exactly, as far as the models allow.

Substitution. Swapping one video model for another means replacing one node and adjusting its settings, while the rest of the graph stays the same.

Designing graphs for a studio

A few principles help graphs serve a studio rather than becoming personal tangles.

Standard templates. Build a small number of template graphs for common shot types: image-to-video with references, video-to-video restyle, keyframe interpolation, upscale and finish. Each shot starts from a template.

Inputs at the edges. Put all shot-specific inputs, such as prompt, frames, references and adapter, in clearly labelled input nodes at the start of the graph. Keep processing nodes generic. This makes templates reusable and makes it obvious what changed between shots.

Standard outputs. End every graph with nodes that save outputs in the project's naming convention, format and location, and write a log entry.

Version the graphs. Store graphs in version control, like code. Record in each shot's log which graph version produced it.

Pin model versions. Record the exact model and adapter versions in the graph. Where possible, keep local copies of open-weight models used in a production, so that a shot can be regenerated even if the public version changes.

Test suites. As Chapter 7 suggested, keep a set of representative test shots. Run them through each template after any change to models or graphs, and compare results.

Kestrel Lane began *The Salt Road* in web interfaces and moved to graphs after the first sequence, when they found themselves repeating the same

twelve steps for every shot. Their template library ended up with seven graphs. The storm sequence, nineteen generated shots, was produced from a single template with a spreadsheet of per-shot inputs.

Agents in the pipeline

An AI agent, in the sense used in this book and in its predecessor *Software Factories,* is a language model in a loop with tools: it receives a goal, decides on actions, calls tools to carry them out, observes the results, and continues until the goal is met or it stops. In software development, agents write and test code. In a synthetic studio, agents can operate generation tools, graphs, editing applications and asset systems.

Several developments in 2025 and 2026 made this practical. The Model Context Protocol, or MCP, an open protocol for connecting AI assistants and agents to tools, gained wide adoption. Generation services began offering MCP servers so that agents could call them directly. Runway, for example, offered an MCP server that let compatible agents generate images and video and, from August 2026 according to its changelog, list, inspect, edit and run saved Workflows. Node-based environments and asset systems offered programming interfaces that agents could use. And general-purpose agents became capable enough to carry out multi-step creative tasks with supervision.

What agents do well in a studio

Batch operations with judgement. Running a template graph across a list of shots, checking each output for obvious failures such as wrong aspect ratio, black frames or missing audio, retrying with a new seed where appropriate, and reporting results.

Preparation. Turning a script breakdown into a shot list, filling shot cards from the production draft, preparing input files and folders, and assembling reference sets from approved images.

Logging and records. Writing generation log entries, updating the shot inventory, producing provenance records, and checking that every shot in the cut has a complete record.

Conversions and deliverables. Transcoding, naming, packaging, generating proxies, producing review files, and checking deliverables against a specification.

First-pass review. Flagging shots with likely problems: identity drift relative to a reference, text in frame, possible logos, colour mismatches against neighbours. An agent cannot judge whether a shot is good. It can find candidates for a person to check.

Research and comparison. Running a test suite across several models and preparing a side-by-side comparison for the director.

What agents should not do

Make creative decisions without review. Selecting takes, approving shots, locking cuts and choosing a look are the director's decisions. An agent can propose; a person disposes.

Act on consent-sensitive material unsupervised. Anything involving a real person's likeness or voice should have a person in the loop at every step.

Spend without limits. Generation costs money per second. An agent retrying a failing shot in a loop can spend a month's budget in an afternoon. Give agents explicit budgets, per task and per day, and have the tools enforce them.

Publish or deliver. Uploading to platforms, sending deliverables, posting publicly and communicating with outside parties should require human approval.

Controls for agents

The lessons of *Software Factories* apply directly. An agent in a synthetic studio should be treated as a capable but untrusted operator:

- **least privilege**: give the agent access only to the tools, projects and folders its task needs;
- **budgets**: cap spending per task and per period, enforced by the service or a proxy, not by the agent's own good behaviour;
- **credentials held outside the agent**: where possible, let a gateway or proxy hold service keys, so the agent never sees them;

- **review gates**: require human approval for creative selection, consent-sensitive actions, spending above thresholds, and anything leaving the studio;
- **logs**: record every action an agent takes, with inputs and outputs, as part of the provenance record;
- **untrusted inputs**: treat content an agent reads, such as web pages, documents and files from outside, as potentially containing instructions that try to redirect it, and do not let such content trigger actions outside the task.

Asset management

Behind graphs and agents sits the studio's asset system: where files live, how they are named, how versions are tracked and how they relate to each other.

Small studios can manage with a disciplined folder structure, a naming convention and a spreadsheet. Larger studios use production tracking and asset management systems from the visual effects world. Either way, the system needs to represent:

- **projects, sequences and shots**, with their status;
- **assets**: characters, locations, props, references, adapters, captures, graphs, look documents;
- **versions** of every asset and shot output;
- **dependencies**: which shot used which version of which assets, made with which graph and models;
- **approvals**: who approved what, and when;
- **consent and rights references** for each asset that needs them.

The dependency record is what makes the system valuable. When a character's reference set changes, it tells you which shots used the old version. When a model is withdrawn, it tells you which shots depend on it. When a distributor asks how a shot was made, it gives the answer.

Review

Review is where a pipeline meets judgement. A studio needs a regular, structured way for the director and department leads to see work, comment on it and approve or reject it.

Good review practice for synthetic work:

- **review in context**: show shots in their sequence, not just alone;
- **review at quality**: full resolution, calibrated display, full speed;
- **compare to references**: have the character bible, look document and previous versions at hand;
- **record decisions**: every approval and rejection logged against the shot version, with notes;
- **limit rounds**: decide in advance how many rounds of revision a shot gets before it is redesigned, cut or accepted.

Many review tools allow frame-accurate comments and side-by-side version comparison, and some integrate with production tracking. Agents can prepare review sessions, assemble comparison reels and transcribe notes into tasks.

An overnight run

To make the role of an agent concrete, consider how Kestrel Lane used one for the conform of the storm sequence, described here as an illustrative workflow.

The task, written by the producer, was: “For every generated shot in sequence 7 of the locked cut, produce the final version at 4K using the recorded inputs, matching the edited timing, and prepare a comparison reel for review.” The agent was given access to the shot inventory, the generation logs, the template graphs, the asset folders for the sequence and the video service through its MCP connection, with a spending cap for the task set at the service, and no access to anything else.

The agent read the inventory and logs, prepared inputs for each shot, and ran the final-quality template for each, using the recorded seeds and settings where the service allowed. For each output, it compared the result against the cut version: duration, framing, and a frame-by-frame

similarity check. Where the final version differed noticeably from the draft, which happened for three shots, it did not try alternatives; it flagged them for review, as instructed. It upscaled the matching shots with the standard upscaling graph, wrote log entries and asset records, and assembled a reel showing draft and final versions side by side, with the flagged shots first.

In the morning, the director reviewed the reel. Two of the flagged shots were acceptable; the third needed a new final generation, which the director made by hand. The whole conform for the sequence took one night of machine time and an hour of human review. Without the agent, the same work had taken the producer two days on the previous sequence. The spending cap was never reached; had a loop of retries begun, it would have stopped the run.

Securing the studio

A synthetic studio holds valuable and sensitive material: unreleased footage, performers' scans and voice recordings, trained adapters of real people, service credentials with spending power, and client material under confidentiality. It deserves the basic security any small business should have, plus a few measures specific to its work.

- **Separate accounts and keys** per service and per purpose, with spending limits set at the service where possible. Do not share a single key across people, projects and agents.
- **Keep credentials out of graphs and project files.** Many workflow files can embed settings, and a shared graph can leak a key. Use the tool's secret storage or environment configuration instead.
- **Restrict access to consent-sensitive assets.** Scans, voice recordings and adapters of real people should be in a separate, access-controlled location, with access logged.
- **Back up and test restores.** Generated material that cannot be regenerated is irreplaceable.
- **Vet custom nodes and plugins.** Node-based tools have large ecosystems of community extensions, which run code on your machines. Install only what you need, from sources you trust, and keep a list.

- **Treat incoming files and pages as untrusted** when agents process them, as described above.

None of this is unusual. It is easy to neglect in a small studio, and the consequences of neglecting it, a leaked performer scan or an emptied service account, are serious.

A reference pipeline

Pulling these elements together, a reference pipeline for a small synthetic studio looks like this:

1. **Development** produces the script, world and character documents, under version control.
2. **Pre-production** produces look documents, reference sets, character bibles, key frames, previs and the production draft, all registered as assets.
3. **Template graphs** are built and tested for the project's shot types.
4. **Production** runs graphs per shot, by hand or by supervised agents, writing outputs and log entries to the asset system. Shot elements are ingested with their camera metadata.
5. **Review** approves shot versions.
6. **Editorial** cuts approved versions, maintaining the shot inventory.
7. **Conform and VFX** produce final versions from recorded inputs, upscale, composite and clean up.
8. **Finishing** grades, mixes and produces deliverables.
9. **QC and provenance** check deliverables and produce provenance and disclosure records from the asset system.
10. **Archive** stores everything needed to regenerate, repair or re-version the film.

This pipeline is deliberately model-agnostic. Models appear only inside graph nodes. If a model disappears, the nodes change and everything else stays.

The next chapter turns to the economics of this pipeline: what it costs, how long it takes, and who does the work.

CHAPTER 17

Cost, Schedule and Crew

Generative tools are often described as making filmmaking cheap. That is partly true and largely misleading. They make certain images far cheaper than they were. They do not make films cheap, because the expensive parts of a film were never only the images: they were the time of skilled people, the iterations needed to get things right, the clearances, the finishing and the delivery. This chapter covers how to budget and schedule a synthetic production honestly, how generation costs actually behave, whether to run models locally or in the cloud, how roles and crews are changing, and what large studios have asked of their production partners.

How generation is priced

Commercial video generation is priced in a few ways, and the method affects how costs behave.

Per second of output. Many APIs charge a price per second of generated video, varying with resolution, model tier and whether audio is included. As examples of published prices in late 2026: Google announced Gemini Omni 1.2 Flash on 5 November 2026 at US$0.07 per second for 720p video, describing this as a 30 per cent reduction from the previous version. A third-party comparison checked on 2 November 2026 listed, among others, Google's Veo 3.1 on Google Cloud at US$0.40 per second with audio and US$0.20 without, and Runway's Gen-4.5 through its API at US$0.12 per second at 720p. Prices change often and differ between tiers and regions; check current pricing pages.

Credits. Subscription products usually sell monthly credits, consumed at different rates by different models, durations and resolutions. Credits make the effective price per second harder to see. Work it out for your common shot types.

Compute. Running open-weight models on your own hardware or rented cloud graphics processors costs electricity, hardware depreciation or rental by the hour. The cost per second depends on the model, the hardware, the resolution and the settings.

Whatever the pricing, three factors multiply the headline price:

1. **The attempt ratio.** You pay for every attempt, not just the take you keep. If a shot takes eight attempts, it costs eight times the per-attempt price.
2. **Duration overhead.** You generate longer than you use, to have handles and to use the middle of the shot.
3. **Pipeline stages.** Drafts, previs, alternative angles, upscaling, video-to-video passes and regenerations at conform all cost money.

An illustrative calculation

To show how these multipliers work, consider a purely illustrative example, not drawn from any real production. Suppose a twelve-minute film contains eighty generated shots averaging five seconds on screen. That is four hundred seconds of generated material in the final cut.

If each shot is generated at eight seconds to allow handles, and takes on average six attempts across drafts and finals, the production generates 80 × 8 × 6 = 3,840 seconds of video. At a hypothetical blended price of US$0.15 per second, that is US$576. Upscaling, video-to-video corrections and alternative coverage might double it. A full cost for generation in this hypothetical case might therefore be in the low thousands of dollars.

That figure is small compared with the cost of shooting the same images conventionally, which is the source of the claim that generation is cheap. But look at what it leaves out. The same eighty shots, at even half an hour of skilled human time each for preparation, review, selection and logging, consume forty hours. Character and look development, editing, sound, music, grading, VFX cleanup, QC, legal review and delivery all take far longer. In most real synthetic productions, as in conventional ones, people's time is the largest cost. Budget it first.

Building a budget

A synthetic production budget follows the structure of a conventional budget, with some lines changed and some added. Typical sections:

Development and writing. Unchanged in principle. Writers' fees, research, development time.

Pre-production. Look development, reference sets, character bibles, previs. Larger than in many conventional productions, because synthetic productions front-load visual decisions.

Performers. For hybrid productions, actors for shot elements, performance capture and voice. Include fees for any digital replica creation or use, as the agreements in Chapter 19 may require.

Capture. Location capture sessions, equipment, permissions.

Production: shot elements. Crew, equipment, locations, stage or LED volume time.

Production: generation. Model subscriptions, API usage, compute, storage, and the people who run it.

Post-production. Editing, VFX, sound, music, grading, finishing, QC.

Rights, legal and insurance. Legal review of model terms and clearances, insurance, consent documentation, music licences.

Delivery and distribution. Deliverables, festival fees, marketing materials.

Contingency. Synthetic productions need a healthy contingency, because model changes, price changes and difficult shots are unpredictable. A model being withdrawn mid-production, as happened to at least one feature project in 2026, can add significant cost.

Local or cloud

Whether to run open-weight models on your own hardware or use commercial services is one of the main infrastructure decisions for a studio. Neither is always better.

Commercial services offer the strongest models in many categories, no hardware to manage, quick access to new releases and predictable per-use pricing. Their disadvantages are cost at volume, dependence on a vendor's terms and continued existence, limits on control and adaptation, and the need to upload material to someone else's systems.

Local and self-hosted open-weight models offer control: you can adapt them with trained adapters, combine them freely in graphs, keep data on your own hardware, and continue to use a version indefinitely. The open-weight video models of 2026, such as LTX-2 and open versions of Alibaba's Wan line, were described by their makers as capable of running on consumer or prosumer graphics hardware. The disadvantages are hardware cost, setup and maintenance effort, and, in many categories, lower quality than the strongest commercial models at the time.

Many studios use both: local models for adapters, previs, drafts and sensitive material, commercial services for final shots where they lead. The pipeline approach in Chapter 16 makes this easy, because graphs can mix local and remote nodes.

Whatever you choose, read the licence of every open-weight model. Open weights do not always mean unrestricted use; some licences restrict commercial use, require attribution or set conditions on scale.

Scheduling

A synthetic production schedule differs from a conventional one in where the time goes.

Pre-production is longer. Look, characters, references and previs take more time, because they determine everything that follows.

Production is less bounded. In a conventional shoot, the shoot ends on a fixed day. Generation can continue indefinitely. Set a production end date for each sequence and keep to it.

Post-production overlaps production. Editing begins early, and the edit drives further generation. The schedule should allow for this overlap deliberately, with the shot lock rule from Chapter 12.

Conform takes longer. Producing final versions of generated shots, upscaling and checking them against the cut is a significant stage.

Rights and QC need time. Legal review, consent documentation, clearance and provenance records should be scheduled, not squeezed in before delivery.

For a short hybrid film like *The Salt Road*, Kestrel Lane's schedule ran to about five months part-time: six weeks of development and pre-production, a one-day shoot and a half-day on an LED wall, about eight weeks of overlapping generation and editing, four weeks of VFX, sound and finishing, and two weeks of QC, legal review and delivery preparation. These figures are illustrative.

Roles and crews

New roles appeared in film and advertising production between 2024 and 2026 around generative tools, with titles that varied widely. The underlying functions are more stable than the titles. A synthetic studio needs people who can:

- **direct**: hold the vision and make creative decisions;
- **produce**: manage budget, schedule, contracts, rights and delivery;
- **design**: develop the look, characters, worlds and key frames;
- **generate**: run models and graphs skilfully, solve shot problems, and know which tool suits which shot;
- **build pipelines**: design graphs, templates, agents, asset systems and automation;
- **capture**: photograph, scan and record real places, objects and performances;
- **edit**: cut the film, and guard continuity;
- **finish**: composite, grade, mix and deliver;
- **clear**: manage consent, rights, provenance and disclosure.

In a two-person studio, each person holds several of these functions. In a larger one, each may be a team. The important point is that the functions exist whether or not someone is assigned to them. A production with no one responsible for consent and rights will find out at delivery.

Displacement and fairness

It would be dishonest to write about crews without acknowledging that generative tools change who gets work. Some tasks that once employed many people, such as storyboard illustration, concept art iteration, rotoscoping, some kinds of animation, background matte painting, temp voice and library music, can now be done faster by fewer people with AI tools. Performers' and writers' unions bargained hard over these questions in 2023 and again in 2026, as Chapter 19 describes, and artists' organisations have raised concerns about training on their work without consent.

A studio can make its own choices about this. Some that Kestrel Lane made, offered as examples rather than rules: they hire human actors for every character with significant dialogue or emotional weight; they commission original music for score and use generated music only for temp and minor cues; they credit and pay the people whose work trains their character adapters, such as the illustrator who designed the original character sheets; and they do not prompt for or train on the styles of identifiable living artists.

What studios ask of partners

Major studios and platforms began publishing guidance for their production partners on generative AI in 2025. Netflix's guidance, published in August 2025 and reported by trade press at the time, illustrates the shape of these requirements. According to those reports and Netflix's partner documentation, partners were expected to tell their Netflix contact about intended uses of generative AI. Low-risk uses, such as ideation and temporary material, were generally acceptable provided outputs did not recreate copyrighted material or identifiable likenesses, the tools did not store, reuse or train on production data, and the work happened in an enterprise-secured environment. Uses involving final deliverables, talent likeness, personal data, third-party intellectual property or work covered by union agreements required written approval in advance.

Other studios, broadcasters and funders have issued their own guidance. In Australia, Screen Australia published AI guiding principles

emphasising transparency, ethical use, protection of practitioners' rights, consent and fair terms, and governance, and asked funding applicants to disclose AI use, which it said was collected for research and reporting rather than for assessing eligibility.

For a studio that wants to work with such partners, the implications are practical: know which tools you use and their data terms; keep generation logs and consent records; separate temporary from final uses; and be ready to describe any generative AI use in writing before it reaches a deliverable.

A one-page budget shape

For a short hybrid film, a budget can be sketched on one page before detailed costing. The proportions below are illustrative only, representing the shape of Kestrel Lane's thinking for *The Salt Road* rather than real figures. They are offered to show where money tends to go, not as benchmarks.

- **People's time** for development, direction, design, generation, editing, finishing and production management: the largest share by far, even when the people are the studio's own founders and the cost is their time.
- **Performers**: a meaningful share in any hybrid film with human actors, including fees for capture and any replica or alteration terms.
- **Shoot and capture**: crew, equipment, location and stage time; modest for a short film with one shoot day.
- **Generation and compute**: subscriptions, API usage and hardware; noticeable, but smaller than most first-time producers expect relative to people's time.
- **Music and sound**: composer, recordings and licences.
- **Rights, legal and insurance**: a small share that cannot be skipped for commercial work.
- **Delivery and festivals**: fees, deliverables and screeners.
- **Contingency**: proportionally larger than in conventional short films, for model and price changes.

The value of sketching the shape early is that it focuses attention on the right questions. If people's time dominates, as it usually does, then the most important savings come from better planning, better templates and fewer wasted iterations, not from finding cheaper models.

Contracts with clients

When a synthetic studio works for clients, its contracts should address AI use explicitly, because clients' assumptions vary widely. Points to cover, with legal advice:

- **Disclosure of methods.** State that generative tools will be used, broadly how, and which tools or categories of tools, so the client can raise any objections or policy requirements before work starts.
- **Ownership and its limits.** Explain that some generated material may not be protected by copyright in some jurisdictions, and what the client will and will not own as a result.
- **Warranties.** Be careful about warranting that material is free of all third-party rights; the law is unsettled. Warrant what you control: that you obtained required consents, did not deliberately imitate protected works, used tools in accordance with their terms, and performed the agreed clearance checks.
- **Client-supplied material.** If the client supplies images, footage, voices or likenesses, require them to warrant that they have the rights and consents needed for the agreed AI use.
- **Data.** State where the client's material will be processed and whether any tool may retain or train on it.
- **Provenance and disclosure.** Agree who is responsible for platform disclosures and for any labelling obligations in the territories where the work will appear.

Clear terms at the start prevent the most common disputes, which arise when a client discovers late in a project that work was made in a way they did not expect.

Pricing your own work

Finally, a word for studios that sell their services. Clients who have read that AI makes video cheap may expect very low prices. The honest response is to price by the value and the full work, not by the generation cost, and to explain what that work includes: development, design, direction, iteration, editing, sound, finishing, clearances, consent, provenance and delivery. Clients who have been burned by cheap generated work that could not be cleared or delivered tend to understand.

The final part of the book addresses the questions behind many of those clearances: rights, labour, provenance and the frontier.

PART VI

Rights, Labour and the Frontier

CHAPTER 18

Rights, Copyright and Clearance

Every film that is sold, financed, insured or distributed has to answer a simple question: who owns it, and is anyone else entitled to stop it or be paid for it? For a conventional film, the answer is assembled from contracts with writers, performers, crew, composers and rights holders, a process called building the chain of title. For a synthetic film, the same process applies, with new and partly unsettled questions about what copyright protects in AI-assisted work, what rights others may have in generated material, and what the terms of the tools themselves allow. This chapter explains those questions as they stood in November 2026, what courts and authorities had said, and how a careful studio can work while important matters remain unresolved. It is not legal advice. The law differs between countries and is changing, and any commercial production should have qualified legal advice.

Two separate questions

Discussions of AI and copyright often blur two different questions.

The output question: is the material a studio makes with AI tools protected by copyright, and who owns it?

The input question: was it lawful to train the models on the material they were trained on, and can the outputs infringe others' rights?

The output question affects what you own. The input question affects what risks you carry. Both matter to a film's chain of title.

What copyright protects in AI-assisted work

Copyright law in most countries protects original works of human authorship. Whether, and how far, it protects material generated by AI depends on how much human creative control is reflected in the result.

The most detailed official guidance is from the United States Copyright Office. Its report on copyrightability, published on 29 January 2025 as Part 2 of its study of copyright and artificial intelligence, concluded, among other things, that:

- existing law is adequate to address these questions without new legislation;
- using AI as a tool to assist human creativity does not affect copyright protection for the output;
- copyright protects original human expression in a work even if the work also includes AI-generated material;
- copyright does not extend to purely AI-generated material, or to material where there is insufficient human control over the expressive elements;
- whether human contributions to AI-generated output are sufficient for authorship must be analysed case by case;
- based on how generally available technology functioned, prompts alone do not provide sufficient control;
- human authors are entitled to copyright in their expression that is perceptible in AI-generated outputs, and in their creative selection, coordination or arrangement of material in the outputs, or creative modifications of the outputs.

In March 2025, the United States Court of Appeals for the D.C. Circuit, in *Thaler v. Perlmutter,* upheld the Copyright Office's refusal to register a work listing an AI system as its author, holding that the Copyright Act requires human authorship.

Other countries approach the question differently. Some have specific provisions for computer-generated works; others apply general originality tests that have not yet been fully tested against generative AI. The Copyright Office's report surveyed international approaches, and productions intended for international distribution should take advice on each major market.

What this means for a film

A film is a compilation of many contributions. In a synthetic film, the human-authored elements typically include:

- the script, treatment and other writing;
- designs made by people, such as hand-drawn or substantially edited character designs, key frames composed and modified by people, and production design;
- performances by human actors, including those driving generated characters;
- photographed and captured material;
- the selection and arrangement of shots in the edit;
- creative modifications of generated material, such as compositing, painting, grading and substantial editing;
- sound design, music and the mix, to the extent made by people.

Purely generated elements, such as a shot produced from a text prompt and used unaltered, may not be protected on their own in the United States. The film as a whole, as a selection and arrangement of material with substantial human contributions, will usually be protected, but the protection may be thinner for some parts than for others. That matters if someone copies a particular generated sequence, or if you want to license individual elements such as a character design.

Strengthening human authorship

This is one more reason for the practices in earlier chapters. Designed key frames, substantially edited reference sets, recorded performances, deliberate compositing and careful editing all increase human authorship. The generation log, the shot inventory and the asset records document it. When the US Copyright Office registers a work containing more than a minimal amount of AI-generated material, it asks applicants to disclose it and to describe the human contribution; good records make that straightforward.

The input question: training and outputs

Whether it is lawful to train generative models on copyrighted works without permission was the subject of many lawsuits and policy processes around the world in 2023 to 2026, and was not settled in November 2026.

A few developments illustrate the range:

United States. The Copyright Office released a pre-publication version of Part 3 of its study, on generative AI training, in May 2025; reports in October 2026 said a final version had still not been published. Numerous lawsuits were pending. In June 2025, Disney and Universal sued the image-generation company Midjourney, and Warner Bros. Discovery later brought a similar claim; according to the docket in November 2026, the case was in discovery with no ruling on the merits.

United Kingdom. In *Getty Images v Stability AI*, the High Court delivered judgment on 4 November 2025. Getty's secondary copyright infringement claim was dismissed, with the court finding that the model weights at issue did not themselves store or reproduce the copyright works; Getty succeeded only on a narrow trade mark claim relating to watermarks in outputs of some earlier model versions, according to the judgment and law firm summaries. The decision was specific to its facts and to the claims that remained at trial; it did not decide whether training in the United Kingdom would infringe.

Australia. In October 2025, the Australian Attorney-General announced that the government would not introduce a text and data mining exception to copyright law for AI training, and would instead explore other options, including licensing.

European Union. EU law includes text and data mining exceptions with an opt-out for rights holders, and the EU AI Act imposes transparency obligations on providers of general-purpose AI models regarding training content.

Licensing as a response

Alongside the litigation, licensing deals between rights holders and AI companies multiplied. The music settlements and licences described in

Chapter 14 were one example. In December 2025, Disney announced a three-year agreement with OpenAI to allow its Sora video tool and image tools to generate short user-prompted content featuring more than two hundred animated, masked and creature characters from Disney, Marvel, Pixar and Star Wars, excluding talent likenesses and voices; Disney also announced a US$1 billion investment in OpenAI. When OpenAI shut down Sora in 2026, reports said the decision had surprised Disney, which illustrates that licensing arrangements depend on the tools they are tied to.

Some tool providers emphasise that their models are trained on licensed or permitted content and offer contractual protection to customers. Adobe described its Firefly models in these terms. As Chapter 12 noted, Adobe's documentation also stated that third-party partner models available in its applications are not covered by its own commercial-safety claim.

Output infringement

Separately from training, a generated output can infringe if it reproduces or closely imitates protected material: a recognisable copyrighted character, a scene from a film, a photograph, a logo, a piece of music. This risk is within a studio's control. Do not prompt for named characters, franchises, films, artists or brands unless you hold a licence; review outputs for unintended resemblance; and remove or replace anything that resembles protected material. Several studio guidance documents, including Netflix's guidance for partners, specifically ask that outputs not recreate copyrighted material.

Model terms

Every model and service comes with terms of use, and for a studio these are as important as any law. Terms typically address:

- **ownership of outputs**: whether you own what you generate, sometimes depending on your plan;
- **permitted uses**: whether commercial use, broadcast, theatrical distribution or advertising are allowed;
- **content restrictions**: what you may not generate;

- **data use**: whether the provider may store your inputs and outputs or use them to train models;
- **indemnities**: whether the provider will defend you against claims that its outputs infringe, and on what conditions;
- **attribution and disclosure**: whether you must credit the tool or label outputs;
- **changes and termination**: how the provider may change the terms or end the service.

Terms differ between providers and between plans of the same provider, and they change. For every production:

1. record which models and services were used, and on which plan;
2. save a dated copy of the terms in force when you generated material;
3. check that the terms permit your intended distribution;
4. note any indemnity and its conditions;
5. check data terms for sensitive material.

For open-weight models, the licence attached to the weights plays the same role and must be checked equally carefully.

Chain of title for a synthetic film

Putting this together, the chain of title for a synthetic film should include, in addition to the usual documents:

- agreements with all writers, performers, designers and crew, including any AI-related terms;
- consent records for any digital replica, voice clone, likeness or performance alteration;
- location and capture permissions;
- licences for music, stock material and other third-party content;
- the list of AI tools and models used, with dated terms and plans;
- the shot inventory showing which material is generated, captured, shot, composited or transformed;
- records of human contribution, such as generation logs, key frames and edit history;

- resemblance and clearance checks on characters and generated material;
- the provenance record described in Chapter 20.

Distributors and insurers increasingly ask for this information. Errors and omissions insurance, which protects producers against claims such as copyright infringement, defamation and violation of privacy, is required by most distributors. Insurers may ask specific questions about AI use; a well-documented production is easier to insure.

Registering a synthetic film

In the United States, registration with the Copyright Office is not required for copyright to exist, but it is required before suing for infringement of a US work and brings other advantages. Since 2023 the Office has asked applicants to disclose AI-generated material that is more than minimal and to describe the human author's contribution, and its January 2025 report confirmed that approach.

For a synthetic film, a registration application should describe the human authorship clearly: the screenplay, the direction, the performances, the designs, the selection and arrangement of shots, the compositing and editing, the sound and the music, as applicable. It should identify the AI-generated material to be excluded from the claim in general terms. The shot inventory and generation logs make this straightforward. Other countries have different systems; many have no registration at all, and protection depends on proving authorship if challenged. In every case, good records are the evidence.

Questions to ask your lawyer

Legal advice is most useful, and least expensive, when the questions are specific. For a synthetic film intended for commercial distribution, a useful brief to a lawyer includes the shot inventory, the list of tools with their terms snapshots, the consent records, and these questions:

1. In each territory where the film will be distributed, what is likely to be protected in the film, and is any element we care about likely to be unprotected?

2. Do the terms of the tools we used permit our intended distribution, and what indemnities, if any, apply?
3. Do any characters, places, objects or music cues present resemblance or infringement risks?
4. Do our consent records meet the requirements of the agreements and laws that apply to our performers?
5. What disclosures are required in each territory and on each platform?
6. What will our errors and omissions insurer want to see, and is anything missing?
7. How should our contracts with distributors describe AI use and allocate risk?

The answers will differ by project and change over time. Asking the same questions on every project makes the answers easier to compare and the advice cheaper to obtain.

A practical risk approach

While the law remains unsettled, a studio can manage risk with a few principles:

1. **Maximise human authorship** in what matters most: story, design, performance, selection and arrangement.
2. **Prefer licensed or permitted tools** for final material where the choice is available, and read their terms.
3. **Never deliberately imitate** protected characters, works, artists or brands without a licence.
4. **Review every output** for resemblance before it enters the film.
5. **Keep records** from the first day.
6. **Disclose honestly** to collaborators, distributors, insurers and audiences.
7. **Take advice** for commercial work, especially for international distribution.

None of this eliminates risk in an unsettled area. It makes the risks visible, reduces them where possible, and puts the production in a defensible position.

The next chapter turns from what is owned to who does the work: labour, consent and ethics.

CHAPTER 19

Labour, Consent and Ethics

Every technology that changed filmmaking also changed who made films and how they were paid. Synchronised sound ended some careers and began others. Television reshaped studio employment. Digital cameras and editing changed the crafts of the cutting room and the laboratory. Generative AI is doing the same, faster and across more crafts at once, and with a new feature: the technology can reproduce the faces, voices and styles of the people whose work it learned from. This chapter covers the agreements that performers and writers in the United States negotiated in 2023 and 2026, laws on digital replicas, the controversy over synthetic performers, the position of crews and artists, and the ethical commitments a synthetic studio can make. It concentrates on the United States because its agreements are the most detailed and most widely referenced, but notes other jurisdictions where relevant. As with Chapter 18, it is not legal advice.

The 2023 strikes

In 2023, both the Writers Guild of America and SAG-AFTRA, the union representing film and television performers in the United States, went on strike against the major studios, represented by the Alliance of Motion Picture and Television Producers. Artificial intelligence was among the central issues in both disputes, alongside pay, residuals from streaming and other matters. The resulting agreements contained the first detailed collectively bargained rules on AI in film and television.

The writers' terms were described in Chapter 4: AI is not a writer, AI output is not literary material, writers cannot be required to use AI, generated material cannot be used to undermine writers' credit or rights, and companies must disclose generated material given to writers.

SAG-AFTRA's 2023 terms on digital replicas

The performers' 2023 TV/Theatrical agreement, according to SAG-AFTRA's published summary and guidance, distinguished two kinds of digital replica.

An **employment-based digital replica** is a replica of a performer's voice or likeness created in connection with their employment on a production, with their physical participation, for the purpose of portraying them in material they did not actually perform. For these:

- the producer must give at least 48 hours' notice before the session to create the replica;
- the performer's consent to creation must be clear and conspicuous, given in a separately signed or initialled statement or separate writing;
- time spent creating the replica is paid;
- use of the replica in the same production requires consent with a reasonably specific description of the intended use, unless the material remains substantially as scripted, performed or recorded;
- the performer is paid for the days they would have worked had they performed the scenes in person, as estimated in good faith;
- use in other productions requires separate consent, obtained at the time of the use, with only a narrow exception for identified multi-part projects;
- consent given during life continues after death unless explicitly limited; for a deceased performer, consent may be given by an authorised representative or, if none can be found, by the union.

An **independently created digital replica** is one made from existing material, such as previous footage or recordings, rather than through a session. Its use also requires the performer's consent with a reasonably specific description of the use, and bargaining over compensation.

The agreement also addressed **digital alteration** of a performer's recorded performance, requiring consent and a specific description, with exceptions, listed in the agreement, for routine post-production changes that do not substantially change the performance. And it

required notice and an opportunity to bargain when producers used **synthetic performers**, meaning generated characters that are not replicas of any identifiable performer, in place of a human performer.

SAG-AFTRA's guidance to members emphasised that a bolded paragraph in a contract is not enough: consent must be separately signed, specific to identified projects, and cannot be blanket.

The 2026 agreements

The 2023 agreements ran until 2026. The WGA ratified a new agreement in April 2026, keeping its AI provisions and adding the notice requirement on licensing writers' material for AI training described in Chapter 4.

SAG-AFTRA's members ratified a new four-year TV/Theatrical agreement in June 2026, with 91.42 per cent voting in favour according to reports. Among its AI provisions, according to SAG-AFTRA's published answers to members' questions and press reports, producers may use synthetic performers only after notice and bargaining with the union, and must show that the synthetic performer provides significant additional value compared with both a human performer and a digital replica. If bargaining fails, the union may seek arbitration, with damages that are not limited to the compensation a human performer would have lost.

The exact text of collective agreements is long and technical, and summaries necessarily simplify. Productions covered by these agreements should consult the full terms and the unions' guidance, which are published on the unions' websites.

Synthetic performers and the Tilly Norwood controversy

In September 2025, a generated "actress" named Tilly Norwood, presented by its creators as a performer seeking representation, drew wide attention and criticism. SAG-AFTRA issued a statement saying that creativity is and should remain human-centred; that the character was not an actor but a computer-generated character trained on the work of countless professional performers without permission or compensation; that it had no life experience to draw from and no emotion; and that

producers signatory to its agreements could not use synthetic performers without complying with notice and bargaining obligations.

The controversy crystallised a distinction that runs through this book. Generated characters within a story, such as a creature, an animated figure, a background crowd or a stylised protagonist in an obviously synthetic world, are one thing. A generated figure presented as a performer, competing for roles that human actors would play, is another. A studio that uses the first should be thoughtful about disclosure and resemblance. A studio that pursues the second should expect union opposition in covered productions and public scrutiny everywhere.

Laws on digital replicas

Alongside collective agreements, legislators addressed digital replicas.

In **Tennessee**, the ELVIS Act, in force from 1 July 2024, extended the state's protection of a person's name and likeness to their voice, and addressed unauthorised AI simulations.

In **California**, AB 2602, in force from 1 January 2025, made unenforceable certain contract provisions allowing a digital replica of a performer's voice or likeness to replace work they would otherwise have performed, unless the intended uses were reasonably specifically described and the performer was represented by a lawyer or a union in negotiating them. A companion law addressed digital replicas of deceased performers.

At the **federal** level in the United States, the NO FAKES Act of 2025, which would create a federal right in a person's voice and visual likeness, was introduced in both houses of Congress in April 2025. According to the Congress.gov record, it had not advanced beyond committee referral as of November 2026.

In the **European Union**, the AI Act's transparency obligations, applicable from August 2026, require disclosure of deepfakes, as Chapter 20 explains. Several European countries have personality rights that protect likeness and voice.

In **Australia** and many other common-law countries, there is no single general right of publicity, and protection of likeness and voice depends

on a mix of laws, such as passing off, consumer protection, defamation and privacy, and on contracts. Union agreements and funding body guidelines fill some of the gap. Screen Australia's AI guiding principles, for example, emphasise consent and fair terms.

The patchwork means that a production distributed internationally must meet the strictest applicable standard. The simplest approach is to adopt a strict standard everywhere: specific, written, informed, compensated consent for any replica, clone or alteration of a real person.

Crews and artists

Performers and writers have had the strongest collective voice on AI, but generative tools affect almost every craft: concept artists, storyboard artists, illustrators, matte painters, animators, rotoscope and paint artists, editors, sound designers, composers, voice actors, dubbing performers, translators and more. Many artists have objected to their work being used to train models without consent or payment, and some have taken legal action.

The effects on employment are real and uneven. Some tasks are being automated; others are being expanded, because cheaper images allow more ambitious projects; new roles are appearing. A studio cannot resolve these structural questions alone. It can decide how it treats the people it works with.

An ethical framework for a synthetic studio

The following commitments, which Kestrel Lane adopted and which this book recommends, are offered as a starting point. They go beyond what the law requires in many places.

1. Consent for every real person. No real person's likeness, voice or performance will be replicated, cloned, altered beyond routine post-production, or used to train a model, without their specific, written, informed consent and fair payment.

2. Performers see their digital selves. Any performer whose likeness or voice is altered or replicated sees the result before it is locked and can object.

3. Humans for human roles. Characters with significant dialogue or emotional weight are performed by human actors, whose performances may then drive digital characters with their consent.

4. No imitation of living artists. The studio does not prompt for, or train adapters on, the identifiable style of living artists without their agreement.

5. Credit and pay contributors. People whose work contributes to the film, including designers whose sheets train character adapters and performers whose recordings drive generated characters, are credited and paid.

6. Prefer consented and licensed tools. Where there is a choice, the studio prefers tools trained on licensed or permitted material, and records what it uses.

7. Disclose honestly. The studio tells collaborators, distributors and audiences how its films were made, in proportion to the use and in line with applicable rules.

8. Protect data. Scans, recordings and trained adapters of real people are stored securely and deleted when the agreed use ends.

9. Do not deceive. The studio does not depict real people saying or doing things they did not say or do in work that could be taken as factual, and does not fabricate real events.

10. Review the commitments. These commitments are reviewed annually, as tools, laws and agreements change.

Working with unions as a small studio

Small studios often assume that union agreements apply only to large productions. Whether an agreement applies depends on the production company's status as a signatory, the performers' membership, the territory and the type of production, and many unions offer agreements designed for low-budget, independent and short-form work. If you plan to hire union performers, contact the relevant union early, explain your project, including its AI elements, and ask which agreement applies and what it requires.

Even when no agreement applies, the union agreements are useful templates. Notice before creating a replica, separately signed specific consent, payment for creation and use, and limits on reuse are fair practices that make performers more willing to work with you, and they protect you if a dispute arises later.

Credits

Credits are where ethics become visible. Film credits list who did what; they are part of how people build careers and how audiences understand a film. Synthetic productions raise new credit questions. How should a performer whose motion drives a generated character be credited? A designer whose sheets trained an adapter? A pipeline engineer whose graphs produced most of the shots?

There is not yet a settled convention. A sensible approach is to credit people for what they did in plain language ("performance capture: [name] as Mara"; "character design: [name]"; "generative pipeline: [name]"), to avoid crediting tools as if they were people, and to add a factual note about AI use, as Chapter 2 suggested. The guild rules mentioned above also affect writing credits in covered productions.

The next chapter turns to how a film documents and discloses its making: provenance, watermarking and the rules of distribution.

CHAPTER 20

Provenance, Disclosure and Distribution

When anyone can generate a convincing moving image of anything, the question "how was this made?" acquires a new importance. Audiences want to know whether what they are seeing happened. Platforms want to label synthetic content. Regulators require disclosure of deepfakes. Distributors, festivals and awards bodies want to understand AI use. And a studio wants to be able to prove what it made and how. This chapter covers the technical standards for recording and verifying provenance, the main watermarking systems, the disclosure rules that took effect in 2026, platform policies, and the expectations of festivals and awards bodies. It ends with a practical provenance and disclosure workflow.

Provenance, watermarking and detection

Three different technologies are often confused.

Provenance metadata records how a piece of content was made and changed, cryptographically signed so that tampering can be detected. It is attached to the file or stored alongside it. It tells you what the signer claims about the content's history, and lets you verify that the claim has not been altered since it was signed.

Watermarking embeds an imperceptible signal in the content itself, in the pixels or audio samples, that a detector can read. A watermark survives some transformations that would strip metadata, such as re-encoding or screen capture, though not all.

Detection tries to determine from the content alone whether it was generated or manipulated, without relying on metadata or watermarks. Detection is a statistical judgement and is error-prone, particularly as generation improves.

A studio's main tools are the first two. Detection is something others may apply to your work.

C2PA and Content Credentials

The Coalition for Content Provenance and Authenticity, or C2PA, publishes an open technical standard for provenance metadata. Content conforming to it carries what the standard calls a manifest, and what its user-facing branding calls **Content Credentials**. Version 2.2 of the specification was published on 1 May 2025.

In simple terms, a C2PA manifest contains:

- **assertions**: statements about the content, such as what actions were applied (created, edited, composited, AI-generated), what ingredients were used (other assets that went into it, each with their own credentials), and other metadata;
- **a claim**, which gathers the assertions;
- **a signature**, made with a certificate belonging to the signer, binding the claim to the content through cryptographic hashes.

When content is edited by a C2PA-aware tool, the tool can add a new manifest that refers to the previous one as an ingredient, building a chain of provenance. A viewer with a C2PA-aware application can check the signature, see the history, and see whether the content has changed since signing.

For a synthetic studio, three features of the standard matter particularly.

Digital source type. C2PA assertions can state that content was produced by a trained algorithm, using standard vocabulary values. Version 2.2 of the specification introduced updated values for this purpose. This is the standard way to record that material is AI-generated.

Composed assets. A film is a composed asset, built from many ingredients. The specification describes how manifests can record ingredients, so a finished film's credentials can, in principle, refer to the credentials of the shots and elements within it.

Video and streaming. The specification supports video files, including the widely used ISO base media file formats, and its user experience

guidance discusses how streaming video, which arrives in segments at different quality levels, can be validated segment by segment as it plays.

In practice, as of 2026, C2PA support was present in a growing number of cameras, creative applications, generation services and platforms, but was not universal, and many workflows stripped metadata at some step, for example during transcoding. A studio that wants its credentials to survive must check every step of its pipeline.

Watermarking

Several AI providers embed watermarks in generated content. Google's SynthID is the best documented. Google introduced it for images and later extended it to audio, text and video, and launched a SynthID Detector portal in May 2025 that could scan uploaded content for SynthID watermarks. According to Google, by late 2026 the detector was available more widely and could check for watermarks from Google and from partner companies that had adopted SynthID. Google also allowed users of its Gemini app to check whether a video had been made with its AI tools, within upload limits.

Other providers use their own watermarks, visible marks or both. Some commercial video services add visible watermarks on lower subscription tiers.

For a studio, watermarks have two implications. Generated material from some services will carry watermarks that platforms or others may detect, which is one reason to disclose AI use honestly rather than hope it goes unnoticed. And removing or attempting to defeat a provider's watermark may breach its terms and, depending on jurisdiction, other rules. Do not do it.

The EU AI Act's transparency rules

The European Union's AI Act contains transparency obligations in its Article 50, which applied from 2 August 2026. Two are directly relevant to filmmaking.

Providers of AI systems that generate synthetic audio, image, video or text must ensure, as far as technically feasible, that outputs are marked

in a machine-readable format and detectable as artificially generated or manipulated. This obligation falls mainly on the companies that make the tools.

Deployers, meaning those who use AI systems, who generate or manipulate image, audio or video content constituting a **deepfake** must disclose that the content has been artificially generated or manipulated. The Act defines a deepfake, broadly, as AI-generated or manipulated content that resembles existing persons, objects, places, entities or events and would falsely appear to a person to be authentic or truthful. Where the content forms part of an evidently artistic, creative, satirical, fictional or analogous work, the obligation is limited to disclosing the existence of such content in an appropriate manner that does not hamper the display or enjoyment of the work.

The European Commission supported a code of practice on marking and labelling AI-generated content, and published an optional common icon for labelling. National guidance on the code described, for example, how disclosures for video could be shown at the start and during deepfake portions, and how audio deepfakes could require audible disclaimers, while emphasising flexibility for evidently artistic and fictional works.

For a fiction film distributed in the EU, the practical implication is that realistic synthetic depictions of real people, places or events need appropriate disclosure, which for most narrative films can be made in ways that do not interfere with the work, such as in credits, descriptions or an opening card. Productions should take advice on how the rules and the code apply to their specific content.

Platform policies

Major platforms have their own disclosure rules. YouTube, for example, requires creators to disclose content that is realistic and made with altered or synthetic media, including depicting real people saying or doing things they did not, altering footage of real events or places, or generating realistic scenes that did not occur. Creators disclose through a setting when uploading, and YouTube shows a label to viewers. According to YouTube's 2026 update, labels became more prominent, appearing on the video for Shorts and below the player for longer videos;

YouTube may apply labels itself when it detects undisclosed synthetic content, and labels derived from its own AI tools or from verified C2PA metadata cannot be removed by creators.

Other platforms have similar rules, which differ in detail and change. Check each platform's current policy before release.

Festivals and awards

Festivals and awards bodies have taken varied positions on AI.

The Academy of Motion Picture Arts and Sciences added language to its rules for the 98th Academy Awards in April 2025 stating that, with regard to generative AI and other digital tools used in making a film, the tools neither help nor harm the chances of achieving a nomination, and that the Academy and each branch will judge the achievement taking into account the degree to which a human was at the heart of the creative authorship.

Some festivals ask entrants to disclose AI use; some have categories specifically for AI-assisted films; some exclude fully generated films from certain categories. Dedicated AI film festivals have grown. Runway has run an annual AI film festival since 2023; its 2026 edition, the fourth, named *A Face Only a Mother Could Love* by Robert Gaudette as its Grand Prix winner, according to published results.

Before submitting, read each festival's rules on AI, disclose as required, and keep the provenance record ready in case programmers ask.

Distribution

Distributors, sales agents and streaming platforms increasingly ask about AI use as part of delivery. Requests may include:

- a description of AI use in the production;
- the list of tools and models used, with their terms;
- consent and release documentation for replicas and alterations;
- evidence of human authorship for chain of title;
- confirmation that outputs do not infringe third-party rights;
- provenance metadata or a provenance report.

The production records described throughout this book, including the generation log, the shot inventory, consent records, model terms and the chain of title documents, answer these requests. Productions that kept them can respond in days. Productions that did not may need weeks, if they can respond at all.

A provenance and disclosure workflow

Kestrel Lane's workflow, offered as a model:

During production. Every generated, captured, transformed and composited shot has a log entry and an asset record, as described in Chapters 7 and 16. Where tools produce C2PA credentials, they are preserved on ingest. Consent and permission documents are linked to the assets they cover.

At picture lock. The shot inventory lists every shot's method, sources and tools.

In finishing. The finishing pipeline is tested to confirm which steps preserve metadata. Where the finishing tools support it, credentials are added to masters with assertions describing the composition of the film, including that it contains AI-generated material.

At delivery. A **provenance report** is produced: a short document summarising how the film was made, the proportion and nature of AI-generated and transformed material, the tools used, consent arrangements, and any disclosures made. It accompanies the chain of title documents.

At release. Disclosures are made appropriate to each destination: a factual line in the credits; platform disclosure settings where content is realistic and synthetic; festival disclosures as required; and EU-appropriate disclosure for any deepfake content.

The disclosure line in the credits of *The Salt Road* read: "This film combines live-action performance with generative AI imagery and sound. Storm, sea and landscape sequences were generated from designs by the filmmakers. Lighting in three shots of Tom Harlan's performance was digitally altered with his consent. A full production

note is available from Kestrel Lane Pictures." (The actor's name here, like the studio, is illustrative.)

Testing the pipeline for credentials

Content Credentials only help if they survive. Before relying on them, run a simple test: take a generated clip that carries credentials, and pass it through every step of your pipeline (ingest, editing, compositing, grading, upscaling, encoding and uploading), checking at each step with a C2PA-aware inspection tool whether the credentials are still present and valid. In many pipelines, credentials are lost at the first transcode or when a clip is placed in an edit and the edit exported.

Where credentials are lost, there are three options. Use tools that preserve or re-sign them. Preserve the original credentials separately, linked to the asset record, so that the provenance report can cite them. Or sign the finished master with your own credentials, declaring its ingredients and that it contains AI-generated material. Many studios will use the second and third together for some time, until end-to-end support improves.

Disclosure language

Disclosures work best when they are short, factual and specific. Vague statements such as "made with AI" tell audiences little, and exaggerated ones can mislead in the other direction by implying that the whole film was generated. Some examples of the forms Kestrel Lane uses, adapted to context:

End credits (narrative short). "This film combines live-action performance with generative AI imagery and sound. [Specific elements] were generated from designs by the filmmakers. [Any alteration of a real performer] was made with [their] consent."

Online description. "Hybrid live-action and AI-generated film. Storm and sea sequences generated with AI from our own designs; all characters performed by actors."

Festival form. A paragraph drawn from the provenance report, stating methods by share, tools used, and consent arrangements, with an offer of the full report.

Fully generated work. "All images and sound in this film were generated with AI tools, directed and edited by [names]. Characters are fictional and do not depict real people."

Commercial or branded content. Follow the platform's and the territory's rules, which can be stricter for advertising, and make the disclosure visible where the content is seen, not only in a description.

Keep the exact wording used for each release in the project record.

Why provenance is worth it

It is tempting to see provenance and disclosure as burdens. They are better seen as investments. A film with a clear provenance record is easier to register for copyright, easier to insure, easier to sell, easier to defend against false claims, and easier to repair and re-version. Honest disclosure builds trust with audiences who are increasingly sceptical of moving images. And a studio known for clear provenance can attract collaborators, performers and partners who care about how their work is used.

The final chapter looks forward: to world models, volumetric cinema and the research that may reshape synthetic filmmaking next.

CHAPTER 21

World Models and the Road Ahead

When the Cinématographe was first shown in 1895, nobody in that Paris basement could have described what cinema would become. Within twenty years it was an industry, an art and a language that its inventors had not imagined. Prediction in this field is humbling. This chapter does not try to predict. It describes the research directions that were visible in late 2026, explains what each might change for filmmakers if it matures, and suggests how a studio can prepare without betting on any one of them. It ends with an argument about what to keep human.

World models

The most discussed frontier in late 2026 was the interactive world model: a generative system that produces an environment in real time, responds to actions such as moving a camera or a character, and keeps the world coherent as it is explored.

Google DeepMind's Genie line is the best documented. Genie 3, announced in August 2025, generated navigable worlds from text prompts in real time at 24 frames per second and 720p, staying largely consistent for a few minutes with visual memory extending back about a minute, according to DeepMind. It also supported "promptable world events," changing the world through text while exploring. DeepMind listed its limitations frankly: a limited range of actions, difficulty simulating multiple interacting agents, inaccurate representation of real locations, poor text rendering, and interaction limited to a few minutes.

Earlier in 2026, Google made a research prototype called Project Genie available to some subscribers in the United States. In October 2026, DeepMind announced Genie 4, which it described as more realistic and natively multimodal, generating music, ambient sound and interactive dialogue along with the world, and drawing on Gemini models for

appearance and reasoning. Users could modify the simulation in real time with natural language, for example changing the weather, moving the camera or introducing a person. DeepMind reported that testers in creative industries were using it as a draft mode for previsualisation, running multiple takes and changing camera angle, lighting, dialogue and action. In November 2026 Google brought Genie 4 to Project Genie for some subscribers and extended Genie 3 sessions to up to three minutes. DeepMind's stated limitations for Genie 4 included drift in visual detail and object permanence over longer sessions and after several complex edits, and continuing research challenges in physics, audio accuracy, legible text, spatial accuracy and long-term consistency. A project with the filmmaker Anand Gandhi was among the examples Google showcased.

What world models might change

If world models mature, several things could follow for filmmaking.

Previs becomes exploration. Instead of building a previs scene, a director describes a world and walks through it, trying camera positions, blocking and light in real time. This was already happening in limited form in 2026.

Coverage becomes a choice made after the fact. If a scene exists as a consistent world rather than as a fixed clip, a director could choose angles and camera moves after the action is defined, much as a virtual camera works in an animated film.

Virtual production without building. A world model running on an LED volume, responding to a tracked camera, could replace hand-built real-time environments for some uses. Resolution, latency and consistency would need to improve substantially first.

Interactive and branching stories. Worlds that respond to viewers could support forms between film and games. This is not new as an idea; world models could make it cheaper to realise.

What to watch

The questions that will determine how quickly world models matter for production are: how long they stay consistent; how precisely they can be

controlled, including camera, characters and timing; what resolution and frame rate they reach; whether they can be grounded in real captured places and specific designed assets; whether their output can be exported in forms that production tools use, such as video with depth, or 3D scenes; and on what terms they can be used commercially.

Longer, more consistent generation

A quieter but more immediately important direction is the steady extension of what video models can do in a single generation. Clip lengths grew from a few seconds in 2024 to fifteen or thirty seconds for several leading models in 2026. Reference handling grew from a single image to many images, plus video and audio. Some models added consistent character voices. Conversational editing allowed incremental changes.

The research behind this includes better compression of video into latent representations, more efficient attention mechanisms, training on longer clips, and hybrid approaches that combine diffusion with autoregressive generation. If the trend continues, the unit of generation may move from the shot to the scene. That would change the director's job from assembling shots to directing scenes, which is closer to conventional filmmaking.

Control as a first-class input

Research on camera control, motion control and structural conditioning, described in Chapter 8, continues to move into products. The direction is toward models that accept a full scene description: camera paths, character positions and poses, timing, light positions and sound cues, in forms that conventional production tools can export. At that point, the boundary between a 3D animation package and a video model becomes thin: the package describes the scene, the model renders it.

Some describe this as generation replacing rendering. A more useful way to see it is that the long-standing pipeline of layout, animation, lighting and rendering acquires a new, learned final stage that can produce photoreal or stylised images from rough inputs.

Volumetric and 4D capture

Chapter 9 described 4D capture of moving performances as Gaussian splats. Research and production efforts in 2026 aimed to make volumetric capture higher in quality, cheaper and easier to edit. Reports in November 2026 described work at Netflix's Eyeline on improving splats for volumetric capture, and a Netflix internship listing investigating splats as a possible streaming format, though that work appeared to be exploratory.

If volumetric capture matures, a performance could be recorded once and photographed later from any angle, with any lens, under adjusted light. That would change cinematography, editing and visual effects profoundly. It would also intensify the consent questions in Chapter 19, since a volumetric capture of a performer is a very complete replica.

Relightable and editable captures

A current limitation of splats and other neural captures is that they bake in the light present during capture. Research on relightable representations, which separate geometry, materials and lighting, aims to let captured scenes be relit as easily as conventional 3D assets. Tools that estimate material properties from video, such as those described in Chapter 13, are an early production form of this idea. If captures become fully relightable, the distinction between captured and built environments will blur.

New audiences, new forms

Most of this book treats the audience as it has been since 1895: people watching a fixed sequence of images on a screen. Several of the directions above point beyond that. Interactive world models, volumetric captures that viewers can move through, and personalised or branching versions of a film each imply a different relationship between filmmaker and viewer, closer in some ways to games, theatre or installation art.

These forms are not new as ideas. What may be new is their cost. If a world or a performance can be captured or generated once and then experienced from many viewpoints, small studios may be able to make

work that previously required game-studio budgets. The craft questions would be new too: how to direct attention without controlling the camera, how to pace a story the viewer can wander away from, how to credit and pay performers whose captured performances are experienced in ways nobody planned.

A studio does not need to move into these forms to benefit from thinking about them. Capturing and storing assets in forms that could support them, such as splats, 3D data and clean separate audio, keeps the option open at little cost.

Agents as crew

Chapter 16 described agents operating pipelines under supervision. The direction of travel is toward agents that handle larger parts of production: preparing all shots for a sequence from a production draft, running and reviewing generations, assembling rough cuts for a director's review, preparing deliverables and records. The lessons of software factories apply: agents should be bounded by explicit permissions and budgets, their actions recorded, and creative and consent-sensitive decisions kept with people.

A studio that already runs its work through graphs, templates and structured records is well placed to adopt agents safely, because agents work best with well-defined tools and clear records. A studio that works through ad hoc sessions will find agents unpredictable.

Provenance by default

The provenance technologies in Chapter 20 were maturing in 2026: more cameras, applications and platforms supported Content Credentials; watermark detection broadened; regulation in the EU required machine-readable marking by providers and disclosure of deepfakes. The direction is toward provenance as a default property of media, attached at capture or generation and maintained through editing. If that happens, the absence of credentials may itself become a signal.

For studios, this argues for building provenance into pipelines now, rather than retrofitting it later.

Research directions to watch

Pulling these together, the research and product directions most worth watching for a synthetic studio, as of late 2026, are:

1. **Scene-length and multi-shot generation** with consistent characters, light and space.
2. **Explicit control interfaces** that accept camera, pose, layout and timing from standard production tools.
3. **Interactive world models** with longer consistency, higher resolution and export to production formats.
4. **Relightable and editable neural captures**, including 4D performance capture.
5. **Open-weight models** competitive with the best commercial ones, which would change the economics of local pipelines and adaptation.
6. **Audio-visual joint generation** with controllable dialogue performance.
7. **Provenance and watermark robustness**, and their integration in professional tools.
8. **Licensing frameworks** for training data, likeness and voice, and the outcome of pending litigation.

A frontier test bench

One practical way to keep up with research without being distracted by it is to maintain a small **frontier test bench**: a fixed set of challenges, drawn from your real work, that you run against each new capability that appears.

Kestrel Lane's bench, after *The Salt Road*, has eight challenges:

1. the keeper's wife at three ages, consistent, from the illustrator's drawings;
2. a thirty-second walk-and-talk between two characters with matched eyelines;
3. the descending crane shot onto the lighthouse gallery, following a given camera path;
4. the lamp-room splat, explored and relit as night with a storm outside;

5. an actor's recorded performance transferred to a stylised character, with the performance intact;
6. a generated sea that matches the look document without retries;
7. a clip carrying valid Content Credentials through the whole finishing pipeline;
8. a world-model exploration of the coast road that can be exported and used as previs.

Each new model, tool or research release is tried on the relevant challenges, and the results are kept in a dated log. Over time, the log shows which capabilities are genuinely improving for the studio's own work, as opposed to which are merely new. It takes a few hours a month, and it replaces a great deal of reading about tools with direct evidence of what they can do.

Preparing without betting

A studio cannot know which of these will matter most, or when. It can prepare in ways that pay off whatever happens:

- **Keep durable assets.** Stories, designs, reference sets, captured frames, recorded performances and documents are useful whatever models come next.
- **Design for replacement.** Keep models behind graph nodes and templates, so new ones can be adopted quickly and withdrawn ones replaced.
- **Capture the real.** Real places, performances and sounds, captured well, will remain valuable inputs for any future model.
- **Record everything.** Provenance records, logs and consent documents become more valuable as rules tighten.
- **Run small experiments.** Try each new capability on a test shot from a current project. Adopt what helps; ignore what does not.
- **Invest in people.** The skills that transfer across tools are story, design, direction, editing, sound, finishing and judgement.

What to keep human

The question that runs under every chapter of this book is what should remain human in a synthetic studio. The tools will continue to improve. Some tasks that people do now will be done by machines. That has been true of every technology in the history of film.

This book's answer is not a list of tasks but a set of responsibilities.

Meaning. Deciding what a film is about, what it says and why it should exist. Models can produce images and words; they do not care whether a film means anything.

Judgement. Choosing between possibilities: this take, not that one; this cut, not that one; this much, and no more. Judgement is where taste and values become visible.

Performance. The expression of human experience through a human body and voice. Synthetic performers can imitate it. Audiences, performers and their unions have made clear that they value the real thing, and the agreements of 2023 and 2026 protect it.

Consent. Deciding whether to use a person's face, voice or work, and asking them. Only people can give consent, and only people can respect it.

Accountability. Standing behind the work: its truthfulness, its fairness, its effect on the people in it and the people who watch it. A model cannot be responsible for anything. Someone must be.

The Lumières built a box that carried images from the world to an audience. A synthetic studio is a larger box, with some of the world made inside it. The job remains what it was in the basement of the Grand Café in December 1895: to show people something worth seeing, honestly, and to take responsibility for what you show.

Appendix A. Templates

These templates collect the working documents used throughout the book in a form ready to adapt. The formats are deliberately plain; the fields are what matter. Replace angle-bracketed values, remove what does not apply, and keep each document under version control with the project. None of them is a legal document. Consent and release forms in particular must be reviewed by a lawyer for your jurisdiction and, where relevant, checked against the union agreements that apply.

A.1 Shot card

One card per shot in the production draft. It expands one line of the shot list into a working record.

```
SHOT CARD
Project:            <title>          Sequence: <nn>    Shot: <nnn>
Version:            <v01>
Status:             planned | in progress | review | approved | final
Method:             shot | generated | captured | composited | transformed
Story purpose:      <what this shot does for the story, one sentence>
Description:        <shot size, angle, subject, action, camera, setting, light, lens>
Duration on screen: <seconds>        Generate/shoot at: <seconds, with handles>
Key frame:          <file, version>  End frame: <file, version or n/a>
References:         <character bible vX; location set vX; prop set vX>
Adapters:           <name, version, base model, or none>
Controls:           <camera path | pose | depth | motion | none>
Sound notes:        <ambience, effects, dialogue, music>
Neighbours:         <previous shot / next shot, for continuity>
Consent refs:       <consent record IDs for any real person, or none>
Template graph:     <name, version>
Budget:             <max attempts / max spend>
Review notes:       <date, reviewer, decision, reason>
```

A.2 Character bible

One per principal character. Version the whole document.

```
CHARACTER BIBLE
Character:          <name>           Version: <v01>    Date: <date>
Role in story:      <one paragraph>
Description:        <age, build, face, hair, skin, eyes, distinguishing marks>
Movement and voice: <posture, gait, gestures; voice description>
Costumes:           <costume A: description + reference set ID; costume B ...>
Reference set:      <folder, version; list of views included>
```

```
Prompt block:       <the exact reusable text used in prompts>
Adapters:           <name, version, base model, training data folder, test grid>
Performed by:       <actor name(s) for body / face / voice, or "synthetic">
Voice:              <recorded actor | synthetic voice + licence |
                    licensed clone + consent ID>
Consent records:    <IDs, or n/a>
Resemblance check:  <date, method, result, reviewer>
Continuity notes:   <injuries, changes over the story, props carried>
Change log:         <version, date, change, reason>
```

A.3 Consent and release checklist for digital replicas and alterations

Use this as a checklist for discussions with performers and their representatives, and for briefing your lawyer. It is not a contract.

```
CONSENT CHECKLIST: DIGITAL REPLICA / VOICE / ALTERATION
Performer:                  <name>        Representative: <name, union/lawyer>
Project(s) identified:      <title(s); no blanket consent>
Type:                       creation of replica | use of replica | voice clone |
                            alteration of recorded performance | performance transfer
Specific description of use:<scenes, lines, nature of change, approximate extent>
Media and territories:      <where it will be shown>
Duration:                   <how long the consent lasts>
Notice given:               <date; at least the notice any agreement requires>
Separate signature:         <yes: signed separately from the general contract>
Payment:                    <for creation; for use; basis of calculation>
Approval:                   <performer sees result before lock: yes/no;
                            how objections are handled>
Data handling:              <where scans, recordings, adapters are stored;
                            who has access>
Deletion:                   <when and how data and models are deleted;
                            confirmation to performer>
Other uses:                 <prohibited unless separately consented>
Death or incapacity:        <who may decide; any limits>
Applicable agreements/laws: <union agreement; state or national law checked>
Signed:                     <date; copies held by both parties>
```

A.4 Generation log entry

One entry per generation that might reach the film. Most fields can be filled automatically by a graph or agent.

```
GENERATION LOG
Shot:             <nnn>    Take: <tnn>    Date/time: <ISO 8601 with zone>
Operator:         <person or agent ID; supervising person>
Service/software:<name>   Plan/licence: <plan or open-weight licence>
Model:            <name, version>   Adapter(s): <name, version>
Graph:            <template name, version>
Inputs:           prompt: <text>
                  start frame: <file, hash>   end frame: <file, hash>
                  references: <files, versions>   control passes: <files>
Settings:         seed <n>; duration <s>; resolution <w x h>; fps <n>;
                  guidance <n>; reference strength <n>; other <...>
Output:           <file name, hash>   Credentials present: yes/no
Cost:             <credits or currency>
Decision:         selected | runner-up | rejected     Reviewer: <name>
```

```
Reason:          <short note>
Terms snapshot:  <ID of dated copy of the service terms>
```

A.5 Shot inventory (at picture lock)

One row per shot in the locked cut.

```
Columns: Shot | In | Out | Duration | Method | Sources (files) |
         Model or camera | Reference versions | Consent IDs |
         Credentials present | Notes
```

A.6 Provenance report

Produced at delivery; one to three pages.

```
PROVENANCE REPORT
Film:                <title>   Version: <cut/version>   Date: <date>
Producer:            <company, contact>
Summary:             <two or three sentences on how the film was made>
Methods by share:    shot <n shots / % runtime>; generated <...>; captured <...>;
                     composited <...>; transformed <...>
AI tools used:       <service/model, version, plan, purpose> (terms snapshots held)
Human authorship:    <script; direction; performances; designs; editing; sound; music>
Real persons:        <performers whose likeness or voice was replicated or altered,
                     with consent record IDs>
Third-party material:<licences for music, stock, locations, captures>
Clearance checks:    <resemblance and output reviews performed; by whom>
Credentials:         <C2PA credentials on masters: yes/no; signer; what is asserted>
Disclosures made:    <credits text; platform settings; festival forms; EU labelling>
Contact:             <for questions>
```

A.7 Delivery checklist

```
DELIVERY CHECKLIST
[ ] Destination spec obtained and read (version, date)
[ ] Frame rate, resolution, aspect ratio, colour space confirmed
[ ] Picture master in working space archived
[ ] Deliverable(s) encoded to spec; checksums recorded
[ ] Loudness measured to spec (report saved); true peak within limit
[ ] Audio stems (dialogue, music, effects) delivered/archived
[ ] Textless versions, subtitles, captions as required
[ ] QC pass: technical, picture, generated-content faults, audio, credits
[ ] Credits include AI disclosure line
[ ] Platform/festival disclosures prepared
[ ] Provenance report attached
[ ] Chain of title documents assembled, including model terms snapshots
[ ] Consent records filed; deletion obligations scheduled
[ ] Content Credentials checked on final files (if used)
[ ] Archive: project files, graphs, logs, references, adapters, captures
```

A.8 Look document skeleton

```
LOOK DOCUMENT
Film: <title>  Version: <v01>
Palette:       <named colours and relationships>
```

```
Light:         <by day / by night / interiors; direction, quality, sources>
Lens:          <typical focal lengths; depth of field>
Camera habits: <static vs moving; height; framing tendencies>
Texture:       <grain, halation, softness; added in finishing or generated>
Aspect ratio:  <ratio; generate at / frame within>
Never:         <things that never appear>
Prompt block:  <the reusable text>
References:    <style reference set ID, version>
```

Appendix B. A Ninety-Day Plan

This plan is for a small team, from one person to a handful, that wants to move from occasional experiments to a working synthetic studio able to deliver a short hybrid film or a commissioned piece to professional standards. It assumes basic filmmaking skills and modest equipment. The phases overlap; the exit criteria matter more than the dates.

Days 1 to 30: finish something

Goal: complete one short film end to end, and learn where your time goes.

- Choose a ninety-second to three-minute project that suits the tools, as described in Chapter 2.
- Write a one-page script; build a look document and one character bible with a reference set.
- Generate, edit, add sound, grade, title and deliver it.
- Keep a generation log from the first shot, even in a spreadsheet.
- Record total time and cost by stage.
- Write the project record: what worked, what failed, what you would change.

Exit criteria: one finished, delivered film with sound and an honest disclosure line; a generation log; a project record with time and cost by stage.

Days 31 to 60: build the foundations

Goal: turn what you learned into repeatable assets and procedures.

- Write your studio's ethical commitments, adapting Chapter 19.

- Adopt the templates in Appendix A: shot card, character bible, consent checklist, generation log, shot inventory, provenance report, delivery checklist.
- Set up an asset structure: folders, naming convention, version rules, and a simple tracking sheet or system.
- Build three to five template graphs in a node-based tool for your most common shot types, with standard inputs and outputs and automatic log entries.
- Assemble a test suite of five to ten representative shots and run two or three candidate models against it. Choose a primary and a second source for each shot type.
- Save dated copies of the terms for every service you use, and read them for commercial use, data and ownership.
- Do one capture session (a room or an object) with permission, train a splat, and use it in a blocking exercise.
- Do one performance session: record an actor (with written consent for a test) and drive a character with it.
- Write a finishing recipe: frame rate, colour pipeline, upscaler, grain, loudness targets.

Exit criteria: templates in use; asset structure in place; at least three tested graphs; a test suite with recorded results; terms snapshots filed; one capture and one performance test completed; a written finishing recipe; written ethical commitments.

Days 61 to 90: run a real production

Goal: make a short hybrid film or a commissioned piece using the full pipeline, and measure it.

- Develop a three- to twelve-minute project with a production draft using method codes.
- Hold lock meetings per sequence; produce key frames and previs before generation.
- Shoot the performance elements; capture at least one location; generate the rest through template graphs.

- Introduce a supervised agent for one bounded task, such as batch final renders or log maintenance, with a spending cap and review gate.
- Lock picture with a shot inventory; conform; finish to the recipe; run QC.
- Produce a provenance report and assemble chain of title documents, including consent records and terms snapshots.
- Deliver to at least one destination with its specification, and make the disclosures it requires.
- Review the production: time and cost by stage against the first project; attempt ratios by shot type; problems by category.

Exit criteria: a delivered hybrid film or commission meeting a real specification; a complete provenance report and chain of title; measured time, cost and attempt ratios; a list of the three improvements that would save the most time on the next project.

After ninety days

Repeat the cycle with larger projects, and revisit the foundations every quarter: models, terms, templates, test suite and ethical commitments. The landscape will have changed. The habits are what let you change with it.

Glossary

Definitions are given in the sense used in this book. Where a term has a formal definition in a standard, agreement or law, the book's description is a simplification; consult the source for the authoritative text.

ACES. The Academy Color Encoding System, a colour management framework for film and television. Version 2.0 was released in April 2025.

Adapter. A small set of additional model weights, usually a LoRA, trained to teach a base model a specific character, style or concept.

Animatic. A timed sequence of storyboard images, usually with temporary sound, used to test story and pacing.

Assembly. The first complete edit of a film, with all scripted shots or placeholders in order.

Attempt ratio. The number of generations made for each generated shot that is used; the generative equivalent of a shooting ratio.

Autoregressive generation. Generating a sequence one step at a time, each step conditioned on what came before.

Banding. Visible steps in smooth gradients caused by low bit depth, compression or heavy grading.

C2PA. The Coalition for Content Provenance and Authenticity, and the open technical standard for provenance metadata it publishes.

Chain of title. The documents establishing who owns the rights in a film and that the producer holds the rights needed to exploit it.

Character bible. A versioned document gathering everything needed to depict a character consistently: description, reference set, prompt block, adapters, voice and consent records.

Composited (method). A shot combining elements from two or more methods, such as a shot performer against a generated background.

Conditioning. Any input that steers a generative model's output: text, images, references, depth, pose, camera paths, audio or video.

Conform. Rebuilding the locked edit from the highest-quality versions of every shot for finishing.

Content Credentials. The user-facing name for C2PA provenance manifests attached to content.

Continuity strip. One frame from each shot containing a character, laid side by side to reveal drift.

Deepfake. In the EU AI Act, broadly, AI-generated or manipulated image, audio or video content that resembles existing persons, objects, places, entities or events and would falsely appear to be authentic or truthful.

Diffusion model. A generative model trained to remove noise step by step, producing new images or video from random noise guided by conditioning.

Diffusion transformer. A diffusion model whose denoising network is a transformer operating on patches of a latent representation.

Digital replica. In the SAG-AFTRA agreements and some laws, a digital representation of a real performer's voice or likeness used to portray them in material they did not actually perform.

Display-referred. Image data encoded for a particular display, with contrast and colour already rendered; most generated footage is display-referred.

Drift. Unintended change in a character's identity, a location's details, light or colour, within or between generated shots.

First-and-last-frame control. Generating a video that begins and ends on specified images.

Frustum. In LED volume production, the region of the LED wall visible to the camera, rendered from the camera's tracked viewpoint.

Gaussian splatting. A 3D scene representation made of many small, semi-transparent ellipsoids, trained from photographs and rendered in real time.

Generated (method). A shot produced by a generative model with no photographed plate.

Generation log. The record of each generation's inputs, model, settings, output and review decision.

Human authorship. The human creative contribution that copyright protects; in the United States, purely AI-generated material is not protected.

ICVFX. In-camera visual effects: capturing final effects in camera, typically with an LED volume showing a real-time environment.

IMF. Interoperable Master Format, a SMPTE standard for delivering masters to platforms and broadcasters.

Key frame. A designed still image fixing the composition, character and look of a shot, often used as a start or end frame for generation.

Latent space. A compressed representation in which many generative models operate, decoded into pixels at the end.

LED volume. A stage enclosed partly or fully by LED panels displaying real-time environments behind and around performers.

LoRA. Low-rank adaptation; a technique for adapting a large model by training a small set of additional weights.

Lock meeting. A meeting at the end of pre-production for a sequence, at which method, key frame, duration and references are agreed for every shot.

LUFS / LKFS. Units of loudness measured according to ITU-R BS.1770.

MCP. Model Context Protocol, an open protocol for connecting AI assistants and agents to tools and data.

Method code. The production method assigned to each shot: shot, generated, captured, composited or transformed.

Motion transfer. Driving a generated or illustrated character with movement and expression from a recorded performance.

NeRF. Neural radiance field; a scene representation in which a neural network returns colour and density for points and directions in space.

Node graph. A workflow represented as connected nodes, each performing one operation; used in compositing and in generative tools such as ComfyUI.

Open weights. A model whose trained weights are published for others to run and, depending on the licence, adapt.

Photogrammetry. Building textured 3D meshes from many overlapping photographs.

Previs. Previsualisation: a rough moving version of a sequence used to plan shots before production.

Production draft. The shooting script with numbered shots, method codes, key frames, references and sound notes.

Provenance report. A short document summarising how a film was made, the tools used, consent arrangements and disclosures, produced at delivery.

Reference conditioning. Supplying images (and sometimes video or audio) of a subject so a model keeps it recognisable in a new shot.

Reference set. A curated, versioned group of images of a character, location, costume or prop used for consistency.

Scene-referred. Image data representing light in the scene rather than an image for a display, typical of camera log and raw footage and of ACES working spaces.

Seed. The number that fixes a generation's random starting noise.

Shot (method). A shot photographed with a camera.

Shot inventory. A list, at picture lock, of every shot with its method, sources, tools, references and consent records.

Synthetic performer. In the SAG-AFTRA agreements, broadly, a digitally created asset that appears to be a natural performer but is not a replica of an identifiable performer.

Synthetic studio. A production system that turns stories into finished, deliverable films using generative models alongside conventional

capture, with defined stages, controls on consent and rights, a record of how every shot was made, and people accountable for the result.

Transformed (method). Photographed footage altered by a model, for example relit, restyled, extended, de-aged or lip-synced.

Video-to-video. Generating a new video from an input video, keeping some properties such as motion and composition while changing others.

Visual dubbing. Altering a performer's lip and facial movement to match dialogue in another language.

Watermark. An imperceptible signal embedded in content that a detector can read, such as Google's SynthID.

World model. A generative model that produces an environment in real time and responds to actions, keeping the world coherent as it is explored.

Notes and References

Sources are listed by chapter in the order they are first relied on. Figures and statements attributed to companies are reported as those companies published them. Product capabilities and prices were checked in November 2026 and will change. Kestrel Lane Pictures, *Low Tide* and *The Salt Road* are composite, illustrative examples; nothing described about them refers to a real company or production. Web addresses were valid when checked.

Preface and Chapter 1

German Patent and Trademark Office (DPMA). "Lumière's cinematograph." Inventions that made history. Patent of 13 February 1895; first commercial public screening, 28 December 1895, Salon Indien, Grand Café, Paris. dpma.de.

Institut Lumière. "Le Cinématographe Lumière." institut-lumiere.org.

Runway. "Introducing Runway Gen-4." March 2025. runway.com/research.

Kling AI. Release notes, January 2025 to November 2026 (Elements; Kling 3.0; Motion Control; native 4K; Kling 4.0). kling.ai/release-note.

Runway. "Introducing Runway Aleph." 25 July 2025.

Lightricks. "LTX-2 is now open source." 6 January 2026; LTX-2 model card, Hugging Face.

OpenAI Help Center. "What to know about the Sora discontinuation." AP and CNBC, 24 March 2026.

Google blog. "Behind ANCESTRA: combining generative AI with live-action filmmaking"; Tribeca Festival 2025 listing.

Reuters, BBC and *The New York Times*, 18 July 2025, on Netflix's use of generative AI in *El Eternauta*.

Bloomberg, 22 May 2026, and *Forbes*, 26 May 2026, on *Critterz*.

Chapter 2

YouTube Help. “Disclosing use of altered or synthetic content.” support.google.com/youtube.

Chapter 3

Ho, J., Jain, A. and Abbeel, P. “Denoising Diffusion Probabilistic Models.” NeurIPS 2020.

Rombach, R. et al. “High-Resolution Image Synthesis with Latent Diffusion Models.” CVPR 2022.

Peebles, W. and Xie, S. “Scalable Diffusion Models with Transformers.” ICCV 2023.

Google. “Gemini Omni 1.2 Flash.” Google blog, 5 November 2026.

Google DeepMind. “Genie 3: A new frontier for world models.” 5 August 2025.

Google DeepMind. “Genie 4: A new generation of world simulation.” 27 October 2026.

Zhang, L., Rao, A. and Agrawala, M. “Adding Conditional Control to Text-to-Image Diffusion Models.” ICCV 2023.

Wang, Z. et al. “MotionCtrl: A Unified and Flexible Motion Controller for Video Generation.” SIGGRAPH 2024.

He, H. et al. “CameraCtrl: Enabling Camera Control for Video Diffusion Models.” ICLR 2025.

Hu, E. J. et al. “LoRA: Low-Rank Adaptation of Large Language Models.” 2021; ICLR 2022.

Ruiz, N. et al. “DreamBooth: Fine Tuning Text-to-Image Diffusion Models for Subject-Driven Generation.” CVPR 2023.

Autodesk. “Using Neural Layer with LoRA for Enhanced Character Consistency in Autodesk Flow Studio.” 23 September 2026.

Chapter 4

US Copyright Office. *Copyright and Artificial Intelligence, Part 2: Copyrightability*. 29 January 2025.

Writers Guild of America. "Artificial Intelligence" (2023 MBA terms) and "Summary of the 2023 WGA MBA." wga.org.

Writers Guild of America. "WGA Ratifies 2026 Minimum Basic Agreement with AMPTP" (24 April 2026) and "2026 MBA Contract Changes FAQ."

Chapter 5

Google DeepMind. "Genie 4." 27 October 2026 (testers' use for previsualisation; limitations).

Chapter 6

Runway Help Center. "Creating with Gen-4 Image References."

Kuaishou Technology. "Kling AI unveils multi-image reference feature." Investor relations release.

SAG-AFTRA. "Summary of 2023 Tentative Successor Agreement" (TV/Theatrical) and "Digital Replicas" guidance. sagaftra.org.

SAG-AFTRA. Statement on synthetic performers, September 2025; *Variety* and *Deadline* coverage.

Tennessee ELVIS Act (effective 1 July 2024); California AB 2602 (effective 1 January 2025).

Chapter 7

LeadWave. "Best AI Video Generators in 2026." 2 November 2026. Third-party comparison, used only as a dated snapshot of the landscape.

Kling AI. "Introducing Kling 4.0." October 2026.

Runway. "Introducing Runway Gen-4.5." 1 December 2025.

Chapter 8

He, H. et al. “CameraCtrl II: Dynamic Scene Exploration via Camera-controlled Video Diffusion Models.” ICCV 2025.

Runway Help Center. “Performance Capture with Act-Two.” July 2025.

Runway. “Introducing Aleph 2.0 and Edit Studio.” Runway news.

Chapter 9

Mildenhall, B. et al. “NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis.” ECCV 2020.

Kerbl, B., Kopanas, G., Leimkühler, T. and Drettakis, G. “3D Gaussian Splatting for Real-Time Radiance Field Rendering.” *ACM Transactions on Graphics* (SIGGRAPH) 2023.

Jawset. Postshot User Guide; radiancefields.com, Postshot platform page.

Blender Foundation. Blender 5.3 release notes, 10 November 2026.

Roettgers, J. “Next up for Hollywood: photorealistic 3D worlds.” *Lowpass*, 19 November 2026.

XangleCS. “Gaussian splat production with 48 cameras and an LED wall.” 2026.

Chapter 10

Industrial Light & Magic. “Groundbreaking LED stage production technology created for hit Lucasfilm series *The Mandalorian*.”

Epic Games. “Forging new paths for filmmakers on *The Mandalorian*.” unrealengine.com.

fxguide. “Art of (LED wall) virtual production sets, part two.”

Chapter 11

Autodesk. Flow Studio product pages.

Flawless AI. “Flawless announces *Watch the Skies*, the first visually dubbed feature to launch on Prime Video in the U.S.”; *Deadline*, October 2026, on the sale of TrueSync and DeepEditor technology.

Metaphysic. “How AI brought the movie *Here* to life”; *Variety*, 2024, Zemeckis on *Here*; SIGGRAPH 2025 talk on the Neural Performance Toolset.

Chapter 12

Adobe. “Meet Firefly Video Model.” 12 February 2025; Adobe Research on Generative Extend in Premiere Pro.

Adobe Help. “Partner models in Adobe products.”

Chapter 13

Meta. “New Segment Anything Models Make it Easier to Detect Objects and Create 3D Reconstructions.” 19 November 2025; SAM 3 repository.

Beeble. “SwitchLight 3.0 is here”; Beeble Studio; Unreal Engine plugin documentation.

Chapter 14

ElevenLabs. Iconic Marketplace; *TechRadar* and *Vox* coverage, 2026.

AP. Universal Music Group and Udio settlement, October 2025. Warner Music Group press releases on Udio and Suno, November 2025; Reuters, 25 November 2025.

ElevenLabs. “ElevenLabs launches Eleven Music” (Business Wire, 5 August 2025); Eleven Music terms; *Billboard* on Merlin and Kobalt agreements.

Chapter 15

Academy Software Foundation / ACES Central. ACES 2.0 documentation and output transforms; release 4 April 2025.

Netflix Partner Help Center. Branded sound mix specification; loudness and true peak guidance; IMF delivery specifications; Dolby Vision HDR mastering guidelines.

European Broadcasting Union. EBU R 128, Loudness normalisation and permitted maximum level of audio signals.

ITU-R BS.1770. Algorithms to measure audio programme loudness and true-peak audio level.

Chapter 16

ComfyUI documentation and changelog, including Wan 3.0 partner nodes (v0.33.4, 24 August 2026).

Runway. Changelog; "Connecting to Runway MCP"; "Building and running Workflows with Agent."

Anthropic. "Introducing the Model Context Protocol." 25 November 2024.

Kavungal, A. *Software Factories.* AWC Field Manuals No. 1. Award Winning Creations, 2026.

Chapter 17

Google. "Gemini Omni 1.2 Flash." 5 November 2026 (pricing).

LeadWave. "Best AI Video Generators in 2026." 2 November 2026 (third-party price snapshot).

The Verge and *Realscreen*, August 2025, on Netflix's generative AI guidance for production partners; Netflix Partner Help Center.

Screen Australia. "AI Guiding Principles" and "AI Transparency Statement."

Chapter 18

US Copyright Office. *Copyright and Artificial Intelligence, Part 2: Copyrightability*. 29 January 2025; Part 3, *Generative AI Training*, pre-publication version, 9 May 2025.

Thaler v. Perlmutter, No. 23-5233 (D.C. Cir. 18 March 2025).

Getty Images (US) Inc v Stability AI Ltd [2025] EWHC 2863 (Ch), 4 November 2025; Reuters coverage.

Disney Enterprises, Inc. et al. v. Midjourney, Inc., No. 2:25-cv-05275 (C.D. Cal.), docket as of November 2026.

The Walt Disney Company. "The Walt Disney Company and OpenAI reach landmark agreement." 11 December 2025; Reuters, same date; Reuters, 24 March 2026.

Attorney-General for Australia. Media release, 26 October 2025; ABC News, 27 October 2025.

US Copyright Office. Copyright Registration Guidance: Works Containing Material Generated by Artificial Intelligence. March 2023.

Chapter 19

SAG-AFTRA. 2023 TV/Theatrical summary; Digital Replicas guidance; "Artificial Intelligence Resources."

SAG-AFTRA. Member FAQ on synthetics under the 2026 TV/Theatrical Agreement; AP and *The Wrap*, June 2026, on ratification.

Congress.gov. S.1367 and H.R.2794, NO FAKES Act of 2025.

Chapter 20

C2PA. *Content Credentials: C2PA Technical Specification*, version 2.2, 1 May 2025; C2PA user experience guidance for implementers, time-based media.

Google. "SynthID Detector" (May 2025) and later updates; Google DeepMind SynthID pages.

European Union. Regulation (EU) 2024/1689 (AI Act), Article 50; AI Act Service Desk; European Commission, Code of Practice on marking and labelling AI-generated content and EU labelling icon; Bundesnetzagentur guidance on the code.

YouTube Help. "Disclosing use of altered or synthetic content"; YouTube Community, "Updates to AI content disclosure and labels" (2026).

Academy of Motion Picture Arts and Sciences. Rules for the 98th Academy Awards, April 2025; AP and BBC coverage.

Runway. AI Festival 2026; results per aifilmhistory.org and hollywood.ai.

Chapter 21

Google DeepMind. “Genie 3” (5 August 2025) and “Genie 4” (27 October 2026). Google blog, “Project Genie” and “Project Genie update featuring Genie 4” (18 November 2026).

Roettgers, J. *Lowpass*, 19 November 2026.

Appendices

Templates in Appendix A are original to this book and may be adapted freely by readers for their own productions.

Index

ACES, 87, 128, 203, 213
Act-Two, 69, 94, 212
adapter, 8, 27, 47, 63, 72, 94, 134, 139
Aleph, 4, 70, 113, 209
Ancestra, 5
animatic, 41, 203
assembly, 104, 203
attempt ratio, 150, 201, 203
audio description, 125
autoregressive, 22, 189, 203
banding, 89, 128, 203
Beeble, 113, 213
Blender, 42, 69, 75, 94, 212
C2PA, 180, 203, 215
camera control, 8, 23, 68, 189, 210
CameraCtrl, 23, 67, 210
captions, 19, 52, 125, 133
chain of title, 161, 183, 201, 203
character bible, 47, 59, 107, 116, 126, 134, 145, 151
Cinématographe, 3, 187, 209
ComfyUI, 5, 140, 206, 214
composited, 33, 46, 62, 74, 76, 83, 104, 115
conditioning, 20, 36, 42, 50, 68, 76, 189, 204
conform, 104, 134, 145, 150, 201, 204
consent, 6, 16, 24, 32, 47, 70, 77, 96
Content Credentials, 180, 191, 204, 215
continuity strip, 53, 103, 204
Copyright Office, 34, 162, 211
coverage, 41, 59, 71, 80, 99, 150, 188, 211
Critterz, 210
DCP, 132
de-ageing, 8, 100
deepfake, 98, 174, 179, 191, 204
depth, 10, 14, 19, 41, 60, 68, 76, 88
diffusion, 4, 19, 189, 204, 210
diffusion transformer, 21, 204
digital human, 91, 93
digital replica, 8, 53, 98, 121, 151, 166, 171, 204
Disney, 164, 214
display-referred, 88, 127, 204
Dolby Vision, 130, 213
DreamBooth, 24, 210
drift, 14, 21, 43, 48, 61, 73, 95, 106
EBU R 128, 213
El Eternauta, 6, 115, 209
ElevenLabs, 121, 213
ELVIS Act, 54, 121, 174, 211
EU AI Act, 164, 204
first-and-last-frame, 66, 72, 204
Flawless, 96, 212
Flow Studio, 24, 51, 94, 115, 210
frustum, 84, 204
Gaussian splat, 8, 36, 42, 68, 79, 86, 190, 212
Gemini Omni, 21, 63, 106, 116, 149, 210
generation log, 12, 26, 62, 72, 104, 134, 142, 155
Genie, 22, 43, 187, 210
Getty Images, 164, 214
grain, 17, 39, 111, 129, 200
guidance, 8, 25, 41, 51, 117, 131, 154, 162
Here, 8, 97, 213
human authorship, 7, 38, 124, 161, 183, 205
ICVFX, 83, 205
IMF, 205, 213
Kestrel Lane, 9, 11, 32, 40, 52, 62, 69, 76
key frame, 26, 41, 52, 65, 73, 109, 140, 153
Kling, 4, 50, 63, 70, 94, 128, 209
latent, 20, 189, 204, 210
LED volume, 8, 74, 76, 84, 151, 188, 204
lip-sync, 12, 73, 96, 120
look document, 39, 64, 107, 129, 144, 193, 199
LoRA, 24, 51, 203, 210

loudness, 17, 125, 127, 200, 205, 213
LTX-2, 5, 51, 64, 152, 209
Lumière, 3, 194, 209
Mandalorian, 83, 212
MCP, 142, 205, 214
Metaphysic, 97, 213
method code, 37, 42, 200, 205
Midjourney, 164, 214
Model Context Protocol, 142, 205, 214
motion transfer, 8, 69, 115, 205
MotionCtrl, 23, 67, 210
Méliès, 3
NeRF, 75, 205, 212
Netflix, 6, 80, 115, 131, 154, 165, 190, 209
NO FAKES Act, 174, 215
node graph, 140, 206
open-weight, 5, 24, 51, 64, 69, 128, 140, 150
performance capture, 24, 47, 70, 90, 93, 151, 177, 192
photogrammetry, 36, 75, 206
Postshot, 78, 212
previs, 23, 42, 61, 68, 76, 83, 105, 140
production draft, 33, 44, 65, 109, 142, 191, 195, 200
provenance, 7, 89, 104, 117, 134, 140, 153, 167
provenance report, 134, 183, 200, 206
reference set, 8, 12, 26, 33, 41, 47, 59, 69
relighting, 6, 55, 79, 88, 111
resemblance, 53, 98, 123, 165, 174
rotoscoping, 6, 111, 154
Runway, 4, 23, 50, 63, 69, 94, 113, 140
SAG-AFTRA, 54, 96, 121, 171, 204, 211
SAM 3, 112, 213
scene-referred, 88, 127, 206
Screen Australia, 154, 175, 214
seed, 18, 20, 62, 142, 206
shot inventory, 110, 126, 134, 142, 163, 184, 200, 206
Sora, 5, 64, 134, 165, 209
StageCraft, 84
storyboard, 4, 14, 26, 37, 39, 60, 105, 154
Suno, 123, 213
SwitchLight, 113, 213
synthetic performer, 54, 171, 194, 206, 211
synthetic studio, 6, 11, 26, 31, 42, 47, 75, 97
SynthID, 181, 207, 215
Thaler, 162, 214
Tilly Norwood, 173
Udio, 123, 213
Unreal Engine, 42, 75, 84, 94, 113, 213
upscaling, 6, 21, 88, 108, 111, 128, 146, 150
video-to-video, 4, 23, 61, 70, 76, 90, 113, 141
visual dubbing, 8, 96, 122, 207
Wan, 63, 140, 152, 214
watermark, 164, 179, 191, 207
WGA, 35, 173, 211
world model, 5, 22, 39, 86, 186, 187, 207, 210
YouTube, 17, 182, 210