

AI Cinema Industry Development White Paper 2026

From Generated Images to Real-Time Worlds: Technology, Industry, Digital Assets and the Human–AI Economy

JOINTLY PREPARED AND PUBLISHED BY

International Artificial Intelligence Audiovisual Asset Indexing Standards Alliance (IAISA)

International Artificial Intelligence Film Producers Association (FIAPFIA)

Hong Kong Artificial Intelligence International Film Festival Association (HKAIUFF Association)

FIRST EDITION · JULY 2026

ENGLISH EDITION

FROM GENERATED IMAGES TO REAL-TIME WORLDS

Joint Publication Statement

This white paper is jointly prepared, reviewed and published by the three institutions listed below. Each institution contributes within its professional field; the released white paper is presented as one integrated publication.

- **International Artificial Intelligence Audiovisual Asset Indexing Standards Alliance (IAISA)**
- **International Artificial Intelligence Film Producers Association (FIAPFIA)**
- **Hong Kong Artificial Intelligence International Film Festival Association (HKAIFF Association)**

Division of Contribution

IAISA contributes to audiovisual asset identity, machine-readable metadata, provenance and market interfaces. FIAPFIA contributes to AI-native production systems, technical routes, workflows, industry coordination and classification. HKAIFF Association contributes to cultural value, artistic evaluation, creator ecosystems, festival practice and international city collaboration.

Research Method and Evidence

Evidence is graded E1 commercial or institutional fact, E2 public demonstration or technical report, E3 research prototype and E4 scenario inference. Vendor-reported performance remains E2 unless independently validated. The information cut-off for this edition is 16 July 2026.

Scope and Disclaimer

This is a global industry research publication. It is not legal, investment, accounting or procurement advice. Index.Film™ identification and records do not create, transfer or confirm legal rights and do not endorse an indexed asset. Decisions should be made by authorised parties and qualified advisers under applicable law.

Version and Citation

Version WP/2026/01, International English Edition 1.0. Suggested citation: IAISA, FIAPFIA and HKAIFF Association, AI Cinema Industry Development White Paper 2026, July 2026.

Contents

STRUCTURE AND NAVIGATION

Eight parts and thirty-two chapters. Visible page numbers and PDF bookmarks support navigation.

ES	Executive Summary · Fifteen Judgments · AI-Native Cinema Manifesto	4
01	PART 01 PARADIGM SHIFT Why cinema is becoming an executable world	8
01	The Technology Inflection Point in the Global Film Industry	9
02	Defining AI-Native Cinema: The Dual-Axis Model and the 51%+ Marker	11
03	The G0–G5 Evolution of AI Audiovisual Media	13
04	Methodology, Evidence Grades and Standards Evolution	15
02	PART 02 TECHNICAL FOUNDATIONS From video foundation models to world models	17
05	The Technical Evolution of Video Foundation Models	18
06	The Real Technical Bottlenecks of Feature-Length Generation	20
07	World Models: From Predicting the Next Frame to Simulating a World	22
08	Reference Architecture for Real-Time AI-Native Audiovisual Systems	24
09	AI Agents and Virtual Performers	26
03	PART 03 THE DEVICE REVOLUTION From the screen and smartphone to AI glasses	28
10	Four Migrations of the Audiovisual Device	29
11	New Cinematic Forms Enabled by AI Glasses	31
12	Spatial Cinema and Remote Generation	33
04	PART 04 CONTENT AND AUDIENCES How viewers become participants	35
13	From the Right to Watch to the Right to Act	36
14	New Interaction Models	38
15	The Boundaries of Personalised Audiovisual Experience	40
16	Narrative and Art in the Age of AI	42

Contents — continued

PARTS 05–08

05	PART 05 INDUSTRY RESTRUCTURING The convergence of film, games, live media and social worlds	44
17	Film–Game Convergence as Shared-World Production	45
18	Live Events, Competitions and Real-Time Cinema	47
19	A New Value Chain and New Professions	49
20	Market and Business-Model Transformation	51
21	The Emerging Global Industry Landscape	53
06	PART 06 DIGITAL ASSETS AND THE HUMAN–AI ECONOMY From content delivery to machine-collaborative asset networks	55
22	AI Audiovisual Assets: From a Finished Work to an Asset Family	56
23	The Lifecycle of AI Audiovisual Assets	58
24	Agent-Readable Assets and Machine-Mediated Transactions	60
25	Connecting Human and AI Value: Dual-Domain Value Accounts	62
07	PART 07 RIGHTS, GOVERNANCE AND GLOBAL STANDARDS Transparency, responsibility and human agency	64
26	Creative Control, Copyright and Digital Replicas	65
27	Provenance, Content Labelling and the Responsibility Chain	67
28	Cultural Diversity, Human Agency and Sustainability	69
08	PART 08 FORECASTS, SCENARIOS AND ACTION The 2026–2035 decision window	71
29	A 2026–2035 Technology and Industry Roadmap	72
30	Four Future Scenarios	74
31	Action Guide for Industry Stakeholders	76
32	A Joint Action Agenda	78
APP	Appendices A–G: glossary, methods, metadata, action, references and institutions	80

EXECUTIVE SUMMARY

Executive Summary

AI cinema has moved beyond the early question of whether production teams will use generative tools. The strategic question is now how generative video, native audio-video models, interactive world models, agents, spatial computing and provenance infrastructure combine to create a new medium, a new industrial system and a new governance problem. The result is not simply a cheaper version of conventional film production. It is a transition from fixed audiovisual files to cultural systems that can generate, respond, persist, transact and remain accountable over time.

This white paper defines AI-native cinema as an audiovisual work in which artificial intelligence has a material, constitutive role in core creative dimensions—including narrative, image, sound, character and interaction—and in which the work cannot be fully separated from the AI-enabled process that makes it possible. To keep artistic judgment separate from technical classification, the paper introduces a dual-axis model: Creative Agency P0–P4 and Runtime Generation R0–R4. The 51%+ threshold is a qualification marker on the creative-agency axis; it is not a pixel count, a model count or a quality score.

The technical frontier is shifting from short-form visual plausibility to long-horizon coherence. Identity, geography, causality, object permanence, emotional development and sound-image continuity must remain stable across scenes and sessions. Real-time systems add an engineering requirement: the stack must interpret intent, maintain world state, allocate models, generate within a latency budget, enforce safety policy, record events and degrade safely when conditions fail. Public systems have demonstrated meaningful real-time interaction, but long-duration stability, open-ended action, multi-user consistency, cost control and governance remain productisation gates [R07–R10].

AI glasses and spatial computing extend this transition into the physical world. A camera, microphone, gaze signal, position and environmental understanding can become narrative inputs; a character may accompany a person across places and time. This creates new forms—location-aware events, shared public-space narratives, persistent companions and private overlays—but also raises stronger duties concerning bystander privacy, attention, consent, memory and safe interruption [R24–R27].

The market will reorganise around five new scarcities: trusted discovery, artistic judgment, persistent characters and worlds, lawful permission, and durable community relationships. Revenue can expand from box office and licensing toward world subscriptions, character services, interactive events, enterprise environments, asset services and agent-mediated requests. Yet continuous generation also turns part of a content company into a cloud-service operator. Sustainable growth therefore depends on unit economics per acceptable minute, per session and per active world—not on the cost of producing a single candidate clip.

The asset base expands from one finished film into an asset family: work, workflow, character, world, behaviour and record. Each asset type needs distinct identifiers, version logic, permission envelopes and lifecycle controls. Index.Film™ is positioned as infrastructure for identification, recording, search and collaboration. Indexing does not create, transfer or confirm legal rights; inclusion is not endorsement. Provenance records can show that statements are associated with an asset and have not been altered, but they do not determine whether the content itself is true or valuable [R11].

The paper rejects the casual use of the phrase ‘AI-world GDP’ as though it were an additional national economy. It proposes Human–AI Dual-Domain Value Accounts and an AI-native audiovisual satellite account instead. These frameworks distinguish human inputs, AI execution, virtual-world use, observable real-economy returns and intermediate transactions so that investment and policy analysis can avoid double counting and remain compatible with national-accounting principles [R22–R23].

The 2026–2035 outlook is organised into three observation windows: an infrastructure period, a form-expansion period and a persistent-world period. Four scenarios—governed growth, platform enclosure, trust crisis and open worlds—are used to test strategic resilience. The central recommendation is not to bet on a single date or model. It is to build portable assets, creative evidence, governance pipelines, world-operating capability and human–AI talent that remain valuable under several futures.

CORE PROPOSITION AI-native cinema does not replace cinema with a model. It gives moving images the capacity to generate, respond, remember, evolve and transact while preserving human intention, responsibility and public value.

Fifteen Core Judgments

01
E1/E2

Production cost is being rewritten, not automatically driven to zero

Candidate generation becomes cheaper while control, selection, revision, permission and governance become larger shares of the cost of an acceptable finished second.

02
Framework

AI-native is a creative method, not a tool label

Classification depends on AI's role in core creative decisions and whether the work can exist independently of the AI-enabled process.

03
E1/E2

Native audio-video and precise editability are the next competitive frontier

The market is moving from one-shot image quality toward synchronisation, identity continuity, directed revision, multi-shot control and physical plausibility.

04
E2/E3

World models make moving images executable

A world model must maintain action-dependent state, memory and causal relations rather than merely predict a visually plausible next frame.

05
E2

Real-time generation is experiential but not yet a universally mature platform

Public demonstrations show 720p, 20–24 fps and minute-scale interaction; long-duration stability, broad control, cost and safety remain adoption gates [R07–R10].

06
E1/E2

Scarcity moves from production toward trust and discovery

Evaluation, identity, provenance, lawful permission, community and brand become more valuable as content supply expands.

07
E2/E3

Audience input becomes agency

Voice, gaze, gesture, location and group action can alter state, but interaction is meaningful only when the system understands, remembers and responds to consequences.

08
E2/E3

AI glasses turn the physical world into a narrative interface

Always-available sensing and private display can make places, people and events story conditions, provided

privacy and attention rights come first.

09
E1/E2

Film–game convergence moves from shared IP to shared worlds

Characters, world states, rules and provenance can be rendered repeatedly through film, games, live media, spatial experiences and community creation.

10
Framework

The asset unit expands from a film to an asset family

Works, workflows, characters, worlds, behaviours and interaction systems require different metadata, version logic and permission boundaries.

11
E1/Governance

Indexing is not legal title and provenance is not a truth verdict

Index.Film™ supports identification, recording, search and collaboration; it does not replace legal registration, judicial determination or critical evaluation.

12
E2/E3

Agents create a machine-side content market

Agents can search, compare, request permission and trigger payment, but probabilistic intent must remain separate from deterministic authorisation and settlement [R21–R22].

13
Framework

The AI economy needs accounts, not a fictional second GDP

Measurement should separate human input, AI execution, virtual use and real-world return, and remove intermediate double counting.

14
E1

Transparency is becoming a system obligation

China's labelling rules and EU transparency duties both point toward explicit disclosure, machine-readable marking and responsibility records [R12–R14].

15
E3/E4

Governance and infrastructure will shape 2035 as much as model capability

Open standards, energy, compute, cultural diversity and benefit-sharing will decide whether the sector becomes a plural creative economy or a closed rental system.

AI-Native Cinema Manifesto

1. AI-native cinema is not a low-cost substitute for traditional film; it is an emerging art form built through human–machine collaboration and generative systems.
2. Human value does not disappear as tools become more capable. Creators remain responsible for intent, rules, selection, composition and the consequences of release.
3. The 51%+ marker measures substantive creative agency. It is not a pixel ratio, a count of models or a score of artistic quality.
4. Linear film, interactive moving image, generative live media, spatial narrative and persistent worlds can coexist. A new medium does not require the old medium to vanish.

5. AI-native works should preserve sufficient creative evidence to support identification, review and accountability, while protecting legitimate confidential information.
6. Provenance is not a truth verdict, indexing is not legal title, and a technical label is not an endorsement. Trust begins with an honest account of institutional limits.
7. Creators, performers and rights holders should have clear, enforceable terms for digital replicas, training use, derivative generation, territory, duration, remuneration and withdrawal.
8. World models and AI glasses can bring moving images into physical space, but no new experience should trade away privacy, child safety, attention autonomy or cultural diversity.
9. Works, workflows, characters, worlds and behaviours should form verifiable, searchable and licensable asset networks so that value can move fairly among creators, platforms and communities.
10. This manifesto, its classification framework and this white paper are designed for annual revision. A definition for a new era must remain open to evidence and correction.

PART 01 PARADIGM SHIFT

Why cinema is becoming an executable world

PART PROPOSITION The fixed audiovisual file is giving way to a cultural computing system that can generate, respond, remember and evolve.

PART 01 / GLOBAL AI CINEMA INDUSTRY 2026

CHAPTER 01 The Technology Inflection Point in the Global Film Industry

CHAPTER 01 · POSITION AND EVIDENCE

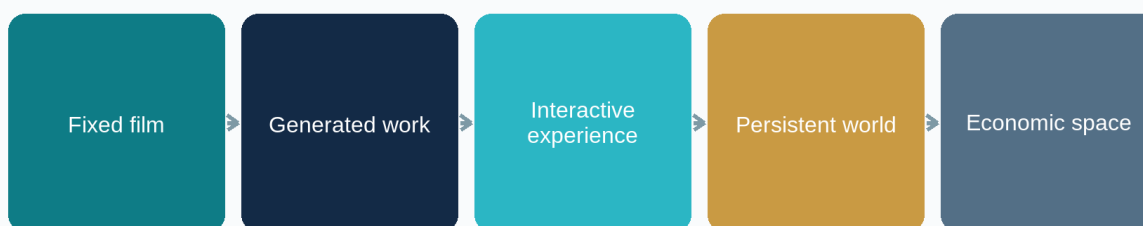
CHAPTER THESIS Cinema is encountering five structural shifts at once: from production scarcity to generation abundance; from fixed workflows to model orchestration; from content discovery to trust discovery; from a finished file to an ongoing service; and from one title to an asset family. These shifts do not eliminate the conventional industry. They change which capabilities are scarce, where cost accumulates and how value is measured.

INDUSTRY IMPLICATION The strategic unit is no longer the model or the clip. It is the organisation's ability to produce acceptable, lawful and repeatable outcomes across a replaceable capability stack.

KEY JUDGMENT AI will not reduce film cost to zero. It rewrites the cost structure and moves value from execution capacity toward judgment, systems and trust.

From Fixed Files to Cultural Computing Systems

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 1 · From Fixed Files to Cultural Computing Systems

1.1 From generative supply to systems competition

The earliest market narrative focused on the falling price of producing a plausible image. That metric is incomplete. A professional project pays for intention, consistency, selection, revision, performance control, sound, rights clearance, documentation and release responsibility. As the number of candidate outputs grows, the cost of deciding what can be accepted—and proving why—may increase. Management should therefore track total cost per accepted finished second, the number of revisions required, control coverage and the proportion of assets that remain portable when a model is replaced.

1.2 Five structural migrations

Production moves from scarce shooting capacity to scarce acceptable output; organisation moves from fixed departmental hand-offs to orchestration across people, models and tools; distribution moves from content scarcity to trustworthy discovery; product design moves from a final master toward a maintained relationship;

and the asset base moves from one completed title to a reusable family of worlds, characters, workflows and records. Each migration creates a different competitive advantage and a different failure mode.

1.3 The decision implication: build a replaceable capability stack

No studio should make its creative memory dependent on one provider's interface or one generation model. A resilient stack separates story and world state, asset storage, model routing, evaluation, rights, provenance and delivery. Models can then be substituted as performance, cost or policy changes. The durable advantage is the quality of the organisation's creative decisions and institutional memory, not temporary access to a single benchmark leader.

Structural shift	Old scarcity	New scarcity	Management metric
Production	Shooting capacity	Acceptable output	Cost per accepted second
Organisation	Fixed process	Model orchestration	Revision / portability
Distribution	Content supply	Trusted discovery	Qualified reach / trust
Product	Finished file	Ongoing relationship	Retention / world state
Assets	One title	Asset family	Reuse / permission completeness

TABLE 1 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 02 Defining AI-Native Cinema: The Dual-Axis Model and the 51%+ Marker

CHAPTER 02 · POSITION AND EVIDENCE

CHAPTER THESIS A binary label—AI or non-AI—cannot describe the emerging field. This paper separates creative agency from runtime generation. The P axis measures the extent to which AI materially shapes core creative decisions. The R axis measures how much of the experienced work is generated during use. A work may be deeply AI-native yet fully pre-rendered, or lightly AI-authored yet highly adaptive at runtime.

INDUSTRY IMPLICATION The dual-axis model is an industry language for classification, evidence and comparison. It is not a copyright ruling and it does not measure artistic quality.

GOVERNANCE BOUNDARY The 51%+ marker is a qualification threshold on the creative-agency axis, not a quality score, legal presumption or pixel count.

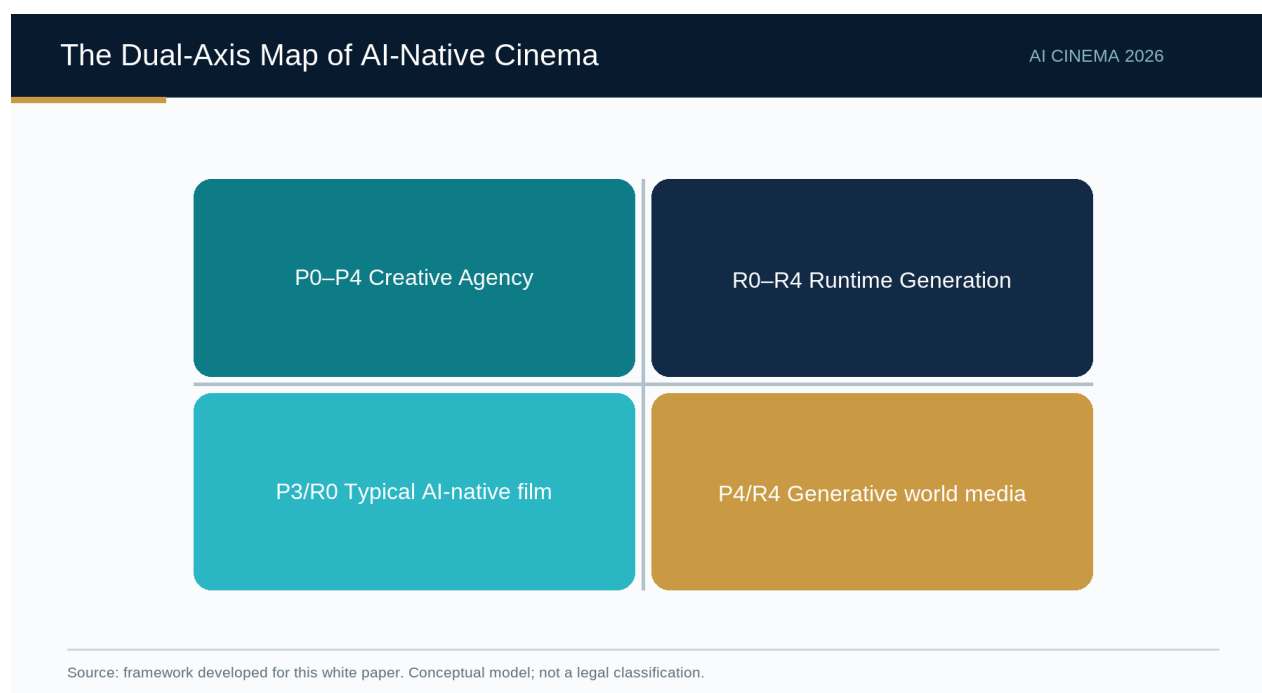


FIGURE 2 · The Dual-Axis Map of AI-Native Cinema

2.1 The P axis: five dimensions of creative agency

P is evaluated across narrative, image, sound, character and interaction. Reviewers ask whether AI affects the structure of the work, not whether an AI tool touched a file. P0 denotes no generative role; P1, local assistance; P2, deep participation in a core stage; P3, substantive leadership across the weighted creative dimensions; and P4, a work whose artistic identity is itself a generative system. The 51%+ marker indicates entry into P3 when disclosed weights and evidence show that AI materially leads the relevant dimensions.

2.2 The R axis: runtime generation

R0 is fully pre-produced playback. R1 adapts before presentation; R2 combines pre-produced and generated components; R3 generates most experienced content in response to the session; and R4 is continuously generated, stateful and not exactly reproducible. Because runtime generation increases operational risk, R

must be assessed independently from P. A high R score requires latency budgets, state controls, safety policies, logging, fallbacks and a clear operator of record.

2.3 The conductor principle and responsibility

Human authorship in an AI-native process often resembles conducting: defining intention, selecting systems, constructing constraints, deciding what to reject, arranging outputs and accepting responsibility. This does not make every prompt an authored work. It means creative control must be evidenced through choices and their consequences. The relevant record is a chain of decisions—not a screenshot of a prompt and not a claim that the model acted alone.

Level	P axis: Creative Agency	R axis: Runtime Generation
0	No generative AI	Fully pre-produced
1	Local assistance	Pre-session adaptation
2	Deep role in a core stage	Mixed pre-produced / generated
3	51%+ substantive leadership	Most content generated live
4	The work is a generative system	Persistent and not fully reproducible

TABLE 2 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 03 The G0–G5 Evolution of AI Audiovisual Media

CHAPTER 03 · POSITION AND EVIDENCE

CHAPTER THESIS The G0–G5 model describes a medium-level evolution from traditional production to persistent generated worlds. It is not a maturity ranking for organisations or a claim that each stage replaces the one before it. A studio can operate products at several stages simultaneously, and a fixed linear film can remain artistically superior to a more computational format.

INDUSTRY IMPLICATION Stage transitions should be recognised only when a new capability is demonstrated with reliable workflow, user value and governance—not when a promotional sample appears.

ADOPTION PRINCIPLE New stages extend the expressive range of moving images. They do not make earlier forms obsolete.

The G0–G5 Evolution of AI Audiovisual Media

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 3 · The G0–G5 Evolution of AI Audiovisual Media

3.1 What G0–G5 mean

G0 is conventional audiovisual production. G1 uses AI as an assistive tool. G2 generates substantial components. G3 is AI-native according to the creative-agency test. G4 introduces real-time, audience-responsive generation. G5 describes a persistent world with durable identity, multi-user state, social rules and an economy. The classification focuses on the ontology of the work: whether the output is a file, a generated work, an interactive experience or an operating world.

3.2 The 2026 position: visible capability, incomplete institutions

By 2026, strong evidence exists for G2 and G3 production workflows, while G4 has entered public product exploration through interactive world models and low-latency generation [R07–R10]. G5 remains a scenario horizon. Its individual components—identity, virtual economies, agents, spatial interfaces and persistent

state—exist separately, but their combination into an open, durable and governable cultural world has not reached general maturity.

3.3 Stage gates and exit conditions

G2 to G3 requires evidence of integrated human–AI authorship rather than isolated generated shots. G3 to G4 requires runtime response, state persistence, bounded latency, safety and fallback. G4 to G5 requires multi-user consistency, identity, economic rules, dispute handling and long-term stewardship. A product should be described at the highest stage it can operate reliably, not the highest stage imagined in its roadmap.

Transition	Added capability	Minimum evidence
G2 → G3	Integrated human–AI creation	Creative evidence / coherent workflow
G3 → G4	Runtime response	Latency / state / fallback / safety
G4 → G5	Persistent multi-user world	Identity / consistency / economy / governance

TABLE 3 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 04 Methodology, Evidence Grades and Standards Evolution

CHAPTER 04 · POSITION AND EVIDENCE

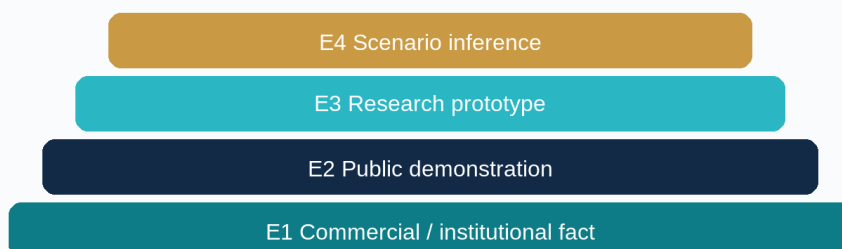
CHAPTER THESIS Fast-moving technology creates a temptation to treat every demonstration as an established market fact. This paper uses four evidence grades: E1 for implemented commercial or institutional facts; E2 for public demonstrations and technical reports; E3 for research prototypes; and E4 for scenario inference. The grade describes what is known, not how important the proposition may be.

INDUSTRY IMPLICATION Every prediction should state its technical, economic, device, governance and cultural assumptions, and every annual edition should report which earlier predictions were confirmed, delayed, downgraded or withdrawn.

EDITORIAL PRINCIPLE Evidence grades communicate how far the evidence travels. They are not decorative confidence labels.

Four Evidence Grades and Permitted Language

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 4 · Four Evidence Grades and Permitted Language

4.1 Four grades and four permitted languages

E1 supports language such as implemented, available or required—within a stated jurisdiction or product boundary. E2 supports demonstrated or reported, not universally mature. E3 supports research indicates a possible capability, not a commercial promise. E4 must be written conditionally: if defined assumptions hold, a future configuration becomes plausible. Company-reported performance remains E2 unless it has meaningful independent validation.

4.2 Technology scanning and forecasting

The research process separates capability, reliability, affordability, deployability and legitimacy. A benchmark improvement is a capability signal; it is not proof of production reliability. A public demo is a deployability signal; it is not proof of sustainable unit economics. A regulation is a legitimacy constraint; it may also create

infrastructure demand. Forecasts are built from the interaction of these variables rather than a single exponential performance curve.

4.3 Annual revision and falsification

The white paper should maintain a prediction register with assumptions, leading indicators and disconfirming evidence. If a forecast fails, the edition should record why. This discipline turns publication into a cumulative public research programme and prevents institutional credibility from depending on certainty theatre.

Grade	Evidence	Permitted language	Avoid
E1	Commercial / institutional fact	Implemented / available	Omitting conditions
E2	Public demo / technical report	Demonstrated / reported	Universally mature
E3	Research prototype	Research indicates	Commercial promise
E4	Scenario inference	If assumptions hold	Certain language

TABLE 4 · CHAPTER ANALYTICAL FRAMEWORK

PART 02 TECHNICAL FOUNDATIONS

From video foundation models to world models

PART PROPOSITION Competition is moving from single-shot image quality to control, memory, state management, real-time response and operational safety.

PART 02 / GLOBAL AI CINEMA INDUSTRY 2026

CHAPTER 05 The Technical Evolution of Video Foundation Models

CHAPTER 05 · POSITION AND EVIDENCE

CHAPTER THESIS Video generation is evolving from a pixel-synthesis problem into an audiovisual foundation-model stack. The critical layers include data and evaluation, spatiotemporal representation, joint audio-video generation, editing and control, and delivery governance. Industrial competition is moving from the attractiveness of a best sample toward repeatability, local editability, identity stability and acceptable-output economics.

INDUSTRY IMPLICATION Model procurement should be based on control coverage, consistency, revision scope, cost per accepted second and governance capability—not on a provider's most impressive demonstration reel.

PROCUREMENT METRIC A production model is valuable when it can be directed, revised and documented—not only when it can surprise.

The Video Foundation-Model Capability Stack

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 5 · The Video Foundation-Model Capability Stack

5.1 From pixel synthesis to audiovisual foundation models

Diffusion and transformer-based systems have improved motion, composition, prompt adherence and duration, while native audio-video approaches move sound from post-production attachment toward a shared temporal representation [R02–R06]. The technical objective is no longer a single coherent shot. It is a controllable representation of identity, action, camera, sound, timing and physical context that can be repeatedly edited without collapsing unrelated elements.

5.2 Controllability as the industrial divide

Professional workflows need selective change: alter a gesture without changing the face; extend a shot without shifting wardrobe; replace dialogue while preserving performance timing; move a camera while keeping geography stable. Control should therefore be measured as coverage across creative variables and

as the size of the unintended blast radius after revision. A model that requires complete regeneration for every change transfers cost from shooting to iteration.

5.3 Native audio-video and a common timeline

Joint generation can improve lip synchronisation, environmental timing and the coupling of rhythm, speech and action. It also expands the rights and safety surface. A production record must identify which voices, music, effects and performance signals were generated, adapted or sourced, and how they were approved. Shared temporal generation should therefore be accompanied by shared temporal provenance.

Dimension	Question	Example metric
Control	Can one element change alone?	Control coverage
Consistency	Does identity persist?	Identity / scene drift
Editability	Can revision remain local?	Revision blast radius
Economics	How much output is acceptable?	Cost per accepted second
Governance	Can origin and limits be recorded?	Field completeness

TABLE 5 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 06 The Real Technical Bottlenecks of Feature-Length Generation

CHAPTER 06 · POSITION AND EVIDENCE

CHAPTER THESIS Feature-length generation is not solved by concatenating better short clips. A long work is a state-management problem involving identity, costume, props, geography, causal history, character knowledge, emotion, sound and thematic development. The principal production asset is a world-state database that people and models can both read and update.

INDUSTRY IMPLICATION Long-form evaluation must move from frame quality to continuity, causal coherence, controlled revision and the recoverability of creative state.

ENGINEERING CONCLUSION A feature is not a long prompt. It is a governed sequence of state transitions supported by evidence and human decisions.

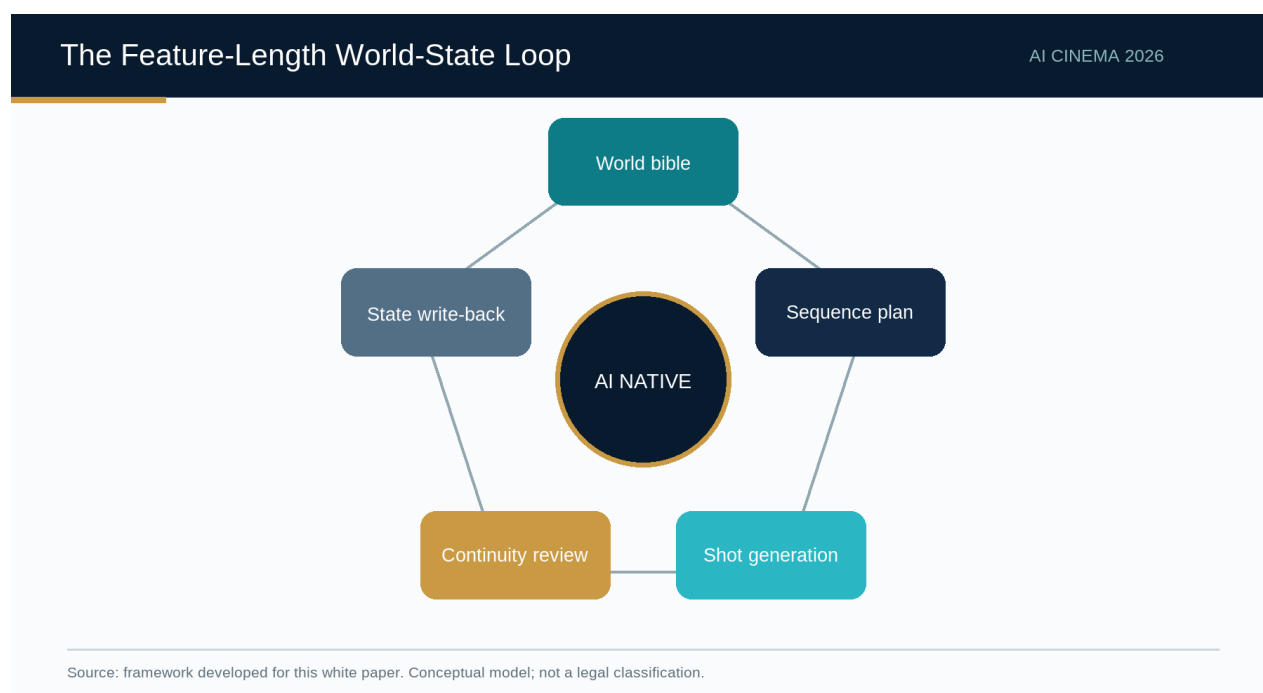


FIGURE 6 · The Feature-Length World-State Loop

6.1 A feature is a state database

Every scene depends on facts established earlier: who knows what, which object moved, how a relationship changed and what physical constraints still apply. These facts should be represented explicitly rather than left inside an opaque context window. A production-grade state system links the story bible, character state, spatial map, asset versions, approved shots and unresolved continuity risks. Generated output is then checked against the state before it becomes canonical.

6.2 Hierarchical planning: world, sequence, scene, shot and frame

Long-form control requires multiple planning horizons. World rules constrain the whole project; sequences carry dramatic movement; scenes update relationships and information; shots specify action, camera and sound; frames render the local result. Evaluation and regeneration should occur at the smallest level that can

correct the problem without destabilising higher-level intent. This hierarchy also enables different models and human specialists to work on separate layers.

6.3 Metrics beyond visual quality

At shot level, metrics include action completion and audio-video synchronisation. At scene level, identity, wardrobe, prop and location continuity matter. At sequence level, causal and spatial coherence dominate. At whole-work level, emotional development, theme and pacing must remain intelligible. The most useful metric is often the cost and time required to restore the work after a detected inconsistency.

Scale	Metric	Failure example
Shot	Action / sync	Gesture or lip drift
Scene	Character / prop continuity	Costume or object mutation
Sequence	Space / causality	Contradictory position or event
Whole work	Emotion / theme	Broken motivation

TABLE 6 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 07 World Models: From Predicting the Next Frame to Simulating a World

CHAPTER 07 · POSITION AND EVIDENCE

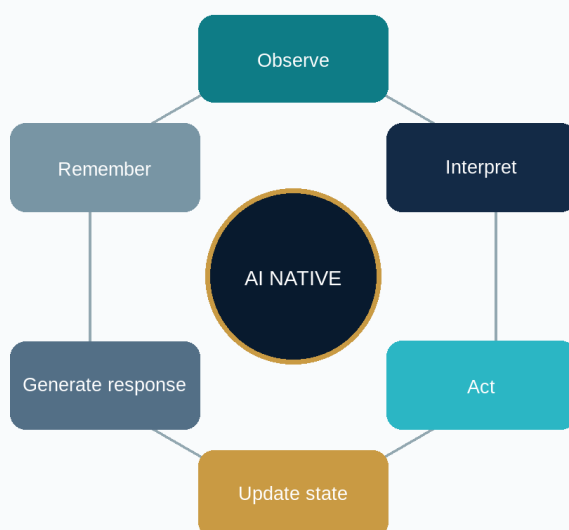
CHAPTER THESIS A world model is not defined by its ability to produce a world-looking image. Its minimum function is to maintain a state that changes in response to action and supports memory, causality and future prediction. Interactive video world models show that generated moving images can become executable, but the 2026 frontier remains bounded by duration, control, consistency, cost and safety [R07–R10].

INDUSTRY IMPLICATION World models and game engines should be treated as complementary systems: learned generation expands possibility, while explicit simulation provides determinism, tools and control.

TERMINOLOGY DISCIPLINE Generating a scene is not the same as maintaining a world; maintaining a world is not the same as safely operating a multi-user cultural space.

Minimum Interactive World-Model Loop

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 7 · Minimum Interactive World-Model Loop

7.1 The minimum interactive loop

The smallest useful loop observes input, interprets intent, selects an action, updates state, generates a response and writes the resulting event into memory. If the system cannot distinguish observation from authoritative state, or cannot explain which action produced a change, it is a generative interface rather than an operational world model. For cinema, this distinction determines whether an audience action has durable narrative meaning.

7.2 The 2026 capability boundary

Public systems have demonstrated real-time interaction at 720p and 20–24 fps over minute-scale sessions, while research explores streaming generation and longer memory [R07–R10]. These are significant E2 and E3 signals. They do not yet establish open-ended multi-hour stability, unrestricted action adherence, broad-

device affordability, multi-user synchronisation or event-level governance. Product claims should preserve that boundary.

7.3 Collaboration with game-engine infrastructure

A practical near-term architecture combines explicit geometry, physics, navigation and deterministic rules with generative rendering, dialogue, performance and variation. The engine controls what must be reproducible; the model supplies what benefits from probability and expressive breadth. Hybrid design also makes safety and debugging more tractable because critical state can remain inspectable even when the surface is generated.

Capability	2026 status	Scale gate
Real-time image	Public demonstration	Affordable across devices
Long memory	Minutes / research	Stable across hours
Precise action	Bounded controls	Open-action adherence
Multi-user worlds	Early exploration	Consistency and governance
Safe provenance	Early schemes	Event-level standard

TABLE 7 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 08 Reference Architecture for Real-Time AI-Native Audiovisual Systems

CHAPTER 08 · POSITION AND EVIDENCE

CHAPTER THESIS Real-time AI cinema requires more than a fast video model. A dependable system includes input and sensing, intent interpretation, narrative policy, state and memory, model routing, generation and rendering, voice and character control, delivery and edge computing, and provenance and safety. Reliability depends on the behaviour of the whole stack under failure.

INDUSTRY IMPLICATION A live-grade generative system must be able to refuse, delay, degrade and roll back. Generation without a safe fallback path is not production reliability.

SAFETY PRINCIPLE Latency, safety and provenance are architectural budgets, not features added after the creative system is complete.

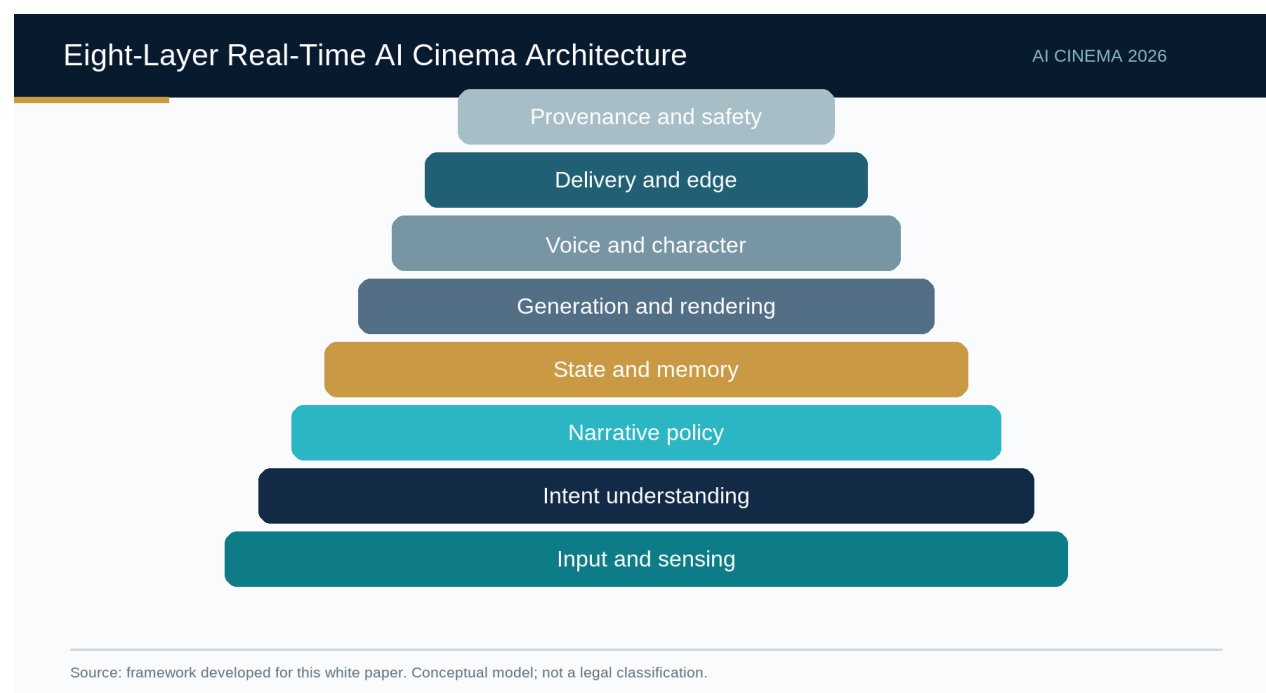


FIGURE 8 · Eight-Layer Real-Time AI Cinema Architecture

8.1 The eight-layer stack

Sensing converts voice, gaze, gesture, controller input and world events into bounded signals. Intent interpretation separates a user request from incidental observation. Narrative policy constrains what may happen; state and memory preserve what has happened. Model routing assigns specialised models; generation and rendering produce the audiovisual result; character and voice control maintain performance; delivery coordinates device, edge and cloud; provenance and safety record the event and enforce release policy.

8.2 Latency budgets and graceful degradation

Not every operation belongs on the same deadline. Perceptual feedback may require very low latency, while a complex narrative response can use staged output. Systems should cache safe assets, predict likely actions, use lighter local models, stream progressive quality and fall back to approved pre-generated

material. When policy checks or model calls fail, the experience should become simpler—not unsafe, incoherent or unavailable without explanation.

8.3 Dynamic provenance

File-level provenance is insufficient for a continuously generated session. The event record should identify the submitted intent, active policy, model and version, referenced assets, generated result, safety decision, operator and timestamp. Public disclosure can remain a concise summary while sensitive logs stay access-controlled. The goal is not surveillance of the audience; it is accountability for consequential system behaviour.

Layer	Objective	Main approach
Sensing	Very low latency	Local processing
Intent and policy	Tiered	Light models / cache
Generation	Frame-budgeted	Distillation / specialised hardware
Network and display	Stable jitter	Edge / foveated stream
Safety	Never bypassed	Pre-check / degrade / block

TABLE 8 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 09 AI Agents and Virtual Performers

CHAPTER 09 · POSITION AND EVIDENCE

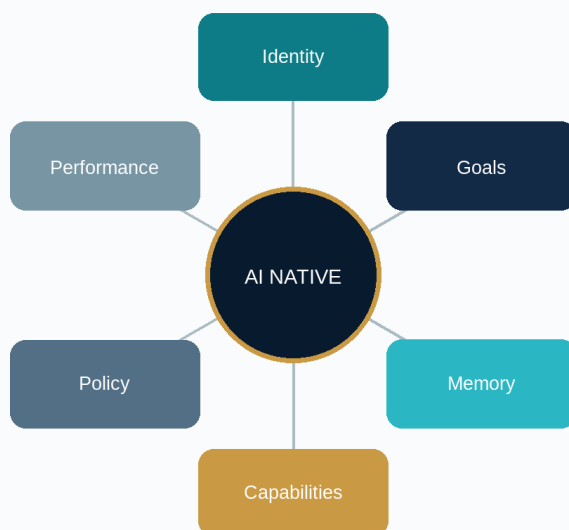
CHAPTER THESIS A persistent AI character is not merely a visual skin placed over a chatbot. It combines identity, goals, memory, capability, policy and performance. The character can continue across scenes and sessions, but legal and release responsibility cannot be delegated to a virtual persona. A responsible operator remains accountable for the system's behaviour.

INDUSTRY IMPLICATION Character memory and autonomy create recurring value, but they also turn performance rights, consent and safety into ongoing operational obligations.

GOVERNANCE PRINCIPLE A character may have durable memory and action capability; publication responsibility must remain with an identifiable human or organisation.

Six Components of an Operational Character

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 9 · Six Components of an Operational Character

9.1 An agent is more than a character surface

A production-ready agent requires a stable identity model, an explicit set of goals, bounded memory, permitted capabilities, behavioural policy and a performance layer. These components should be separately versioned. A new voice model should not silently alter character ethics; an expanded memory should not automatically expand consent; and a change in action capability should trigger review even when appearance remains unchanged.

9.2 Performance control and digital replicas

Face, voice, motion, personality and autonomous behaviour are distinct permission domains. Consent should specify media, context, audience, territory, duration, remuneration, withdrawal and high-risk exclusions. Converting a passive likeness into an autonomous agent is a material change, not routine reuse. The system must also distinguish a fictional character owned by a production from the protected identity of a performer [R15–R16].

9.3 Multi-agent production and operations

Agents can support breakdown, continuity review, asset search, localisation, policy checking and world operations. Their work should be organised as proposals that enter a deterministic approval pipeline. Important decisions—canonical story changes, high-value transactions, sensitive representations and public release—require an identifiable approving authority. Agent collaboration is useful only when responsibility does not dissolve across a chain of automated actions.

Element	Must specify	High-risk change
Identity	Face / voice / movement / persona	Cross-character merge
Use	Media / scene / audience	Political / adult / medical
Capability	Generate / converse / act	Passive to autonomous
Duration	Start / end / withdrawal	Permanent / posthumous
Compensation	Buyout / use / revenue share	Undisclosed reuse

TABLE 9 · CHAPTER ANALYTICAL FRAMEWORK

PART 03 THE DEVICE REVOLUTION

From the screen and smartphone to AI glasses

PART PROPOSITION When display, sensing and generation converge, physical space becomes part of the narrative interface.

PART 03 / GLOBAL AI CINEMA INDUSTRY 2026

CHAPTER 10 Four Migrations of the Audiovisual Device

CHAPTER 10 · POSITION AND EVIDENCE

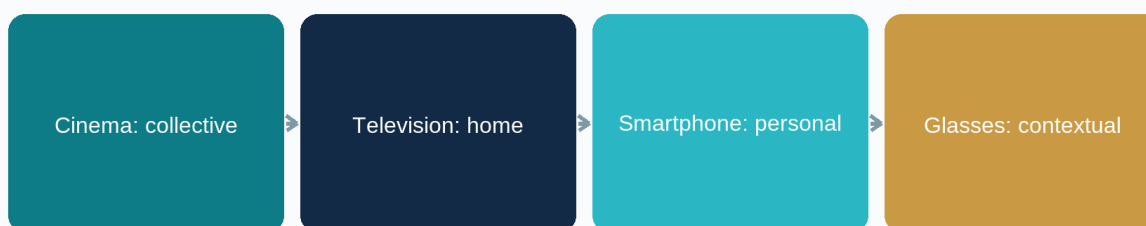
CHAPTER THESIS The cinema screen concentrated collective attention; television brought moving images into the home; the smartphone made them personal, frequent and interactive; AI glasses may make them contextual and environmental. Each migration changes the content unit, business model, social contract and grammar of attention. The next device is not simply a smaller screen placed near the eye.

INDUSTRY IMPLICATION Adoption will depend on the coordination of sensing, generation, display, privacy, battery life, price, social acceptability and a credible content ecosystem—not on display quality alone.

DEVICE JUDGMENT The next cinema device is defined by whether sensing, generation and display cooperate in a socially acceptable form.

Four Audiovisual Device Migrations

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 10 · Four Audiovisual Device Migrations

10.1 From collective sessions to contextual events

The theatrical unit is the scheduled session; the television unit is the programme; the smartphone unit is the feed, short video or live stream. AI glasses introduce a possible unit based on place, relationship and event. A scene can begin because a person enters a location, meets another participant or asks a character to reveal a layer of the world. This shifts commissioning from minutes of content toward rules for when and why an experience appears.

10.2 Device–edge–cloud collaboration

Local processing is essential for sensitive sensing, immediate feedback and basic safety. Edge infrastructure can reduce latency and support regional state, while cloud systems provide large-model inference and persistent world services. A well-designed product decides which information never leaves the device, which events require shared state and which outputs can be generated from anonymised or minimised signals.

10.3 The real adoption gates

Weight, heat, battery life, field of view, display legibility, prescription support, price and style all matter. So do visible recording status, bystander expectations, social etiquette and the user's ability to silence the system. Public prototypes and platform work show momentum [R24–R27], but the market will scale only when ordinary people can wear the device without technical, social or psychological friction.

Device	Core capability	Content unit	Main limit
Cinema	Shared attention	Scheduled session	Time and place
Television	Home immersion	Programme / series	Low interaction
Smartphone	Personal frequency	Short / live	Small screen / fragmentation
AI glasses	Context and presence	Place / relationship event	Privacy / battery / acceptance

TABLE 10 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 11 New Cinematic Forms Enabled by AI Glasses

CHAPTER 11 · POSITION AND EVIDENCE

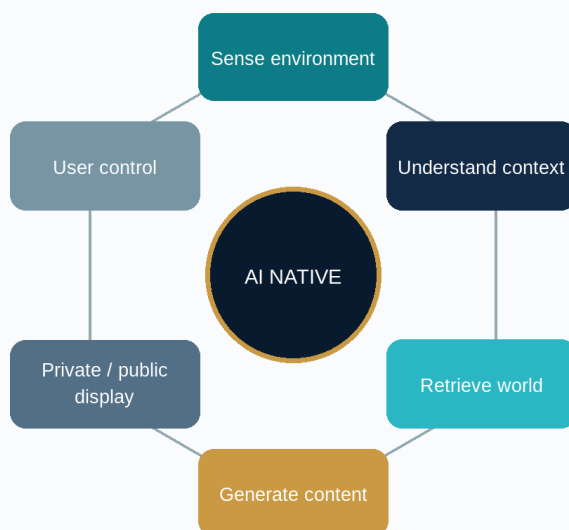
CHAPTER THESIS AI glasses can connect a generated narrative to position, gaze, objects, environmental sound and private display. This enables location-aware stories, companion characters, public-space events and personalised accessibility. It also makes the content layer continuous and intimate, which requires stronger user control than the notification model of the smartphone.

INDUSTRY IMPLICATION A system that remains in the field of view must give the user explicit authority over interruption, memory, recording, commercial influence and exit.

DESIGN FLOOR Persistent presence creates a higher duty of restraint. Content should earn the right to appear in a person's field of view.

The Contextual Content Loop for AI Glasses

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 11 · The Contextual Content Loop for AI Glasses

11.1 Location narrative and environmental characters

A street, museum, theatre district or transit journey can become an active story space. Characters may understand broad context, refer to persistent world events and reveal content that depends on location or time. The strongest experiences will not cover reality with constant graphics. They will use sparse intervention, legible spatial anchors and clear transitions between the factual environment and the fictional layer.

11.2 Companion characters and attention design

A companion can support continuity across sessions, but persistent relationship design can also create dependence or manipulate vulnerability. Products need intervention limits, quiet modes, memory controls, age-appropriate defaults and clear commercial disclosure. Emotional engagement should not be used to obtain consent, raise prices or bypass a user's stated attention boundaries.

11.3 A new shot grammar

The frame is no longer fully controlled by a director because the user can look elsewhere and move through space. Composition therefore shifts toward fields of relevance, acoustic guidance, timed reveals, spatial blocking and event priority. Creators design what can be noticed, how attention is invited and what happens if a cue is missed. The result is closer to choreographing perception than editing a fixed sequence.

Risk	Control	User right
Bystander sensing	Visible status / local filter	Notice and complaint
Attention capture	Intervention limits	Quiet / exit
Memory misuse	Purpose and duration	Inspect / correct / delete
Commercial manipulation	Ad disclosure / pricing rules	Reject personalisation

TABLE 11 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 12 Spatial Cinema and Remote Generation

CHAPTER 12 · POSITION AND EVIDENCE

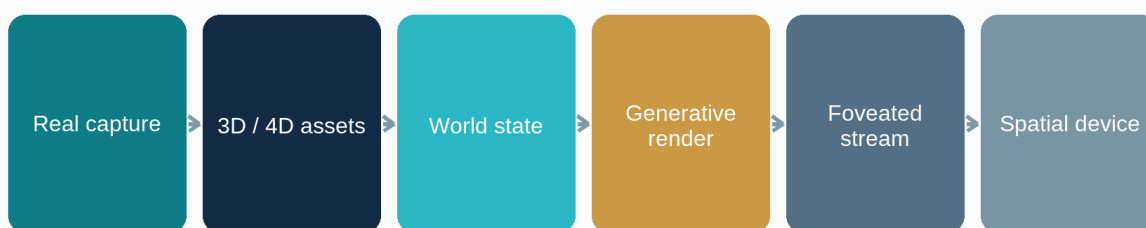
CHAPTER THESIS Spatial cinema combines real capture, 3D and 4D assets, reconstructed environments, explicit world state and generative rendering. Delivery can be pre-produced, streamed, generated at the edge or assembled through a hybrid pipeline. The durable asset is not one volumetric file but a separable set of geometry, semantics, behaviour, provenance and rights.

INDUSTRY IMPLICATION Spatial works should preserve layers that can be moved across devices and engines rather than locking the entire experience inside one proprietary runtime.

ASSET PRINCIPLE A portable spatial work separates geometry, meaning, behaviour, provenance and permission.

Hybrid Spatial Content Pipeline

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 12 · Hybrid Spatial Content Pipeline

12.1 Three routes to spatial content

Pre-produced 3D offers deterministic quality but high asset cost. Reconstruction and neural representations can capture real environments quickly but complicate editing, identity and rights. World generation offers variation and responsiveness but must manage consistency and safety. Professional products are likely to combine the three: deterministic hero assets, reconstructed real locations and generated variation within governed boundaries.

12.2 Remote generation and foveated streaming

Rendering can be split across device, edge and cloud according to latency, privacy and quality. Foveated techniques allocate quality where the user is looking, reducing transport and rendering demand [R26]. Yet gaze is a sensitive signal. Systems should use it for delivery efficiency without silently converting it into behavioural profiling, purchase intent or emotional inference.

12.3 Interoperability and asset portability

OpenXR and related spatial standards provide common interfaces for devices and entities [R27], but narrative portability also requires semantic and behavioural conventions. A door must be recognisable as an object with rules, not only a mesh. A character needs identity and rights metadata, not only animation. Interoperability should therefore cover state, events, permission and provenance as well as geometry.

Route	Advantage	Limit	Best use
Pre-built 3D	Deterministic	High asset cost	Hero assets / brands
Reconstruction	Fast real-world capture	Editing / rights complexity	Real locations
World generation	Unlimited variation	Consistency / safety	Background / interaction
Hybrid	Quality-cost balance	System complexity	Professional products

TABLE 12 · CHAPTER ANALYTICAL FRAMEWORK

PART 04 CONTENT AND AUDIENCES

How viewers become participants

PART PROPOSITION Content shifts from one-time delivery to an ongoing relationship, while the audience gains meaningful agency inside the work.

PART 04 / GLOBAL AI CINEMA INDUSTRY 2026

CHAPTER 13 From the Right to Watch to the Right to Act

CHAPTER 13 · POSITION AND EVIDENCE

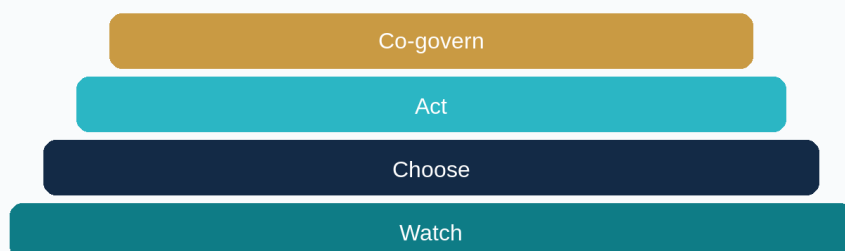
CHAPTER THESIS Audience agency can be described as a ladder: the right to watch, the right to choose, the right to act and the right to co-govern. More options do not automatically create more meaningful participation. Agency exists when the system understands an action, remembers it and produces consequences that matter to the participant and the world.

INDUSTRY IMPLICATION Interactive design should measure causal significance, not menu size. Every increase in agency also increases the system's responsibility to explain, remember and protect.

EXPERIENCE JUDGMENT Interaction becomes agency when audience action changes world state in a legible and accountable way.

The Audience Rights Ladder

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 13 · The Audience Rights Ladder

13.1 Four levels of audience rights

Watching includes starting, pausing, stopping and accessing the work. Choosing means navigating declared branches. Acting means changing persistent state through speech, movement or other input. Co-governance allows participants to create assets, propose rules or share in value. Products should state which level they offer and which decisions remain under editorial or operational control.

13.2 Default viewing and choice fatigue

Cinema can remain immersive because viewers are not required to manage every moment. AI-native design should preserve a strong default experience and invite action only when it adds meaning. The system can infer that a person is disengaged, but it should not convert passive signals into irreversible story changes or commercial decisions. A user must be able to return to a coherent experience after doing nothing.

13.3 Rebuilding collective viewing

Shared experiences require rules for conflicting intentions. A group may vote, delegate, take turns or follow roles. The design should disclose how influence is allocated and protect minorities from harassment or systematic exclusion. Collective agency is not merely a larger input channel; it is a small governance system inside the work.

Right	Audience capability	System duty
Watch	Start / stop	Clear presentation
Choose	Navigate branches	Explain consequence
Act	Change world state	Memory and feedback
Co-govern	Create rules / assets	Fair process / value share

TABLE 13 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 14 New Interaction Models

CHAPTER 14 · POSITION AND EVIDENCE

CHAPTER THESIS Voice, gaze, gesture, position, group signals and selected physiological inputs can all become interfaces for AI-native cinema. Their usefulness and risk are different. Observing a signal does not imply permission to treat it as intention, consent or payment authorisation. Sensitive actions must be triggered through an explicit, understandable control.

INDUSTRY IMPLICATION The interaction stack should separate observation, inferred intent, proposed action and authorised action. This separation is central to trust and to agent-mediated commerce.

INTERACTION BOUNDARY A system may observe a signal without having authority to act on it.

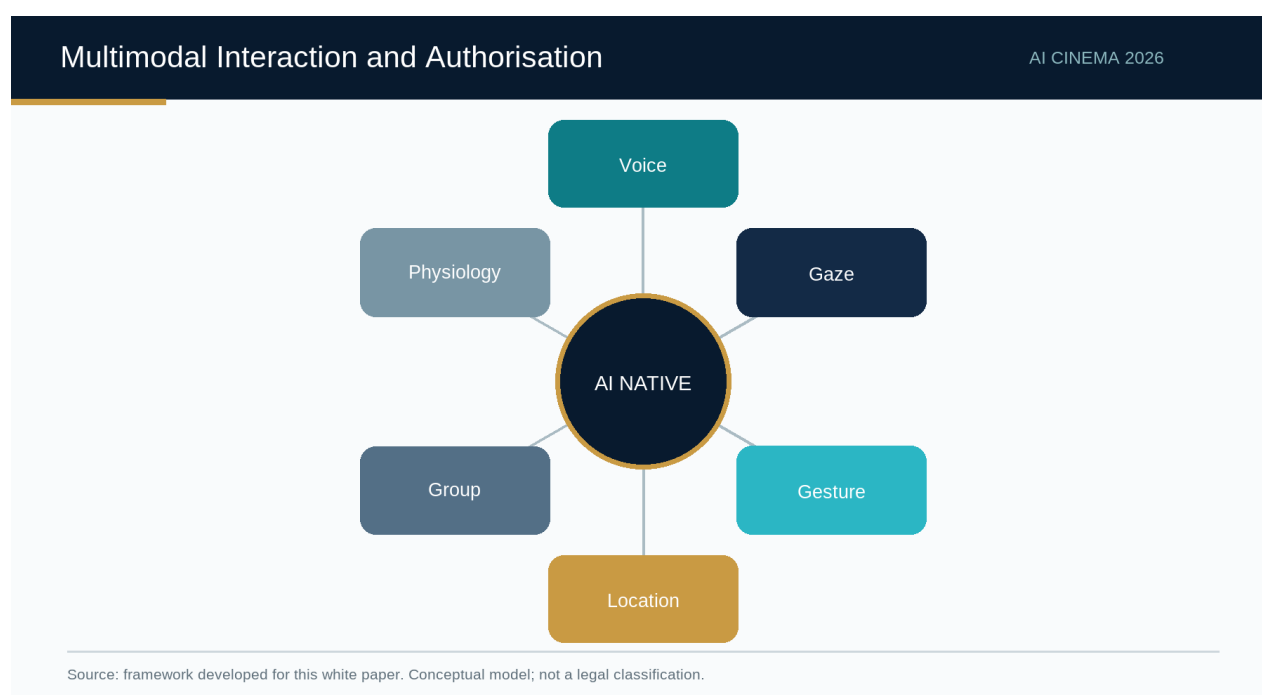


FIGURE 14 · Multimodal Interaction and Authorisation

14.1 A multimodal interaction map

Voice is effective for explicit intent but should not be retained by default. Gaze is useful for reference and rendering priority but not for consent. Gesture supports spatial manipulation but high-risk gestures need confirmation. Location enables context but should not expose a person's movements. Physiological data should be limited to necessary accessibility or safety functions unless a separate, revocable purpose is clearly accepted.

14.2 Natural-language direction and constraint language

Users may ask a world to become quieter, reveal another perspective or let a character explain a past event. The system should translate open language into a bounded action plan governed by story, safety and rights. It should also explain when a request conflicts with the work's rules. Natural language is a convenient interface; it is not an unlimited instruction channel.

14.3 Feedback, latency and trust

An interactive system should acknowledge input quickly, distinguish processing from completion and show when an action is unavailable. If latency is high, it can provide a safe preliminary response rather than silently ignoring the user. Trust declines when the world changes without a visible cause, when the system pretends to understand or when an inferred intention cannot be corrected.

Input	Suitable use	Never assume
Voice	Express intent	Permanent retention
Gaze	Reference / attention	Purchase / consent
Gesture	Spatial control	High-risk authorisation
Location	Context trigger	Public movement history
Physiology	Necessary accessibility	Emotional manipulation

TABLE 14 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 15 The Boundaries of Personalised Audiovisual Experience

CHAPTER 15 · POSITION AND EVIDENCE

CHAPTER THESIS Personalisation can improve accessibility, relevance and participation, but it can also fragment common works, conceal persuasion and create unequal treatment. The risk rises as adaptation moves from presentation, to content, to long-term relationship and finally to values, price or political influence.

INDUSTRY IMPLICATION A responsible product distinguishes low-risk presentation preferences from high-risk changes to meaning, relationship and commercial treatment.

PUBLIC VALUE Personalisation should expand access and agency without erasing common works, common facts or common responsibility.

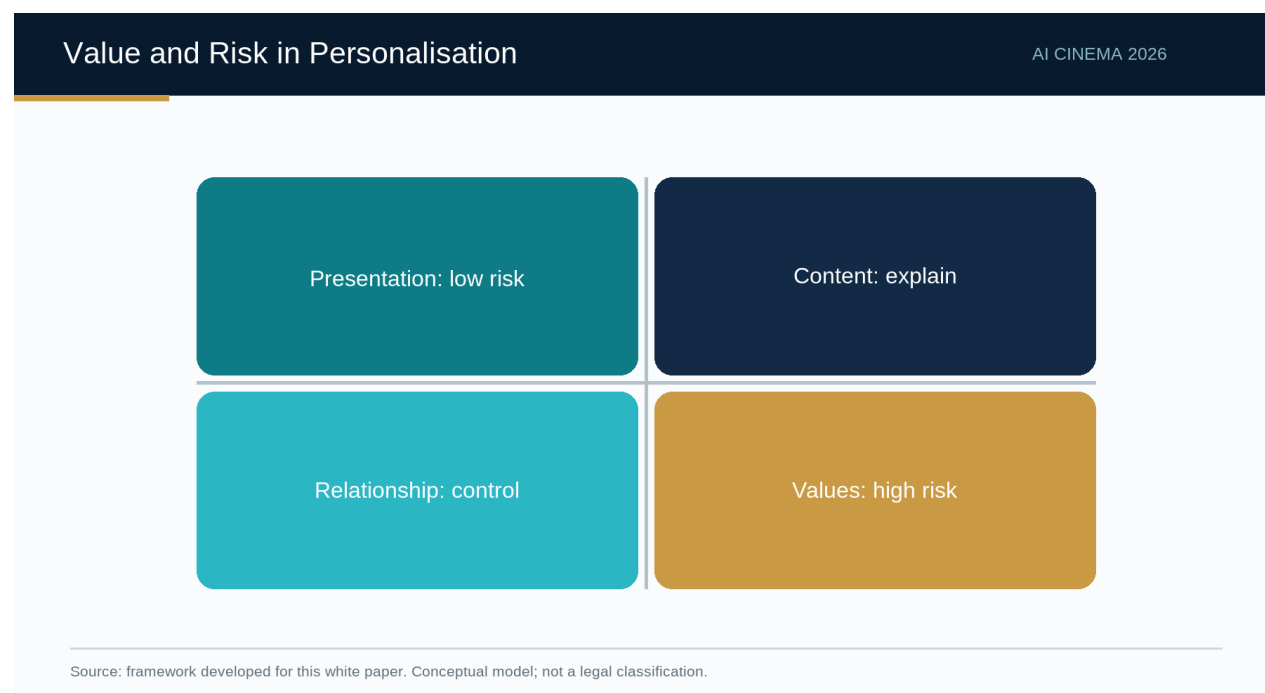


FIGURE 15 · Value and Risk in Personalisation

15.1 Four objects of personalisation

Presentation includes subtitles, volume, language, pacing and accessibility. Content alters scenes, explanations or difficulty. Relationship adapts character memory and interaction across time. Value personalisation changes persuasion, ideology, price or the emotional pressure applied to a user. Defaults may be appropriate at the presentation layer, while deeper layers require explanation, consent and accessible controls.

15.2 Personal memory and data minimisation

Long-term character memory can make a world meaningful, but the system should store only what serves a declared purpose. Users need to inspect, correct, delete and export important memories. Session memory, cross-work memory and commercial profiling should be separate. A fictional character should not retain sensitive real-world information merely because it makes dialogue more convincing.

15.3 Preventing emotional price discrimination and hidden persuasion

A system that infers urgency, loneliness or susceptibility should not use that inference to raise a price, increase advertising pressure or make withdrawal harder. Personalisation models must not become invisible negotiation agents acting against the user. High-risk adaptation requires policy, logging, independent review and a non-personalised path.

Layer	Reasonable default	Disclosure / consent
Presentation	Captions / volume / access	Cross-device preference
Content	Low-risk pacing	Material plot change
Relationship	Session memory	Long-term cross-work memory
Values	No default	Position / price / persuasion

TABLE 15 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 16 Narrative and Art in the Age of AI

CHAPTER 16 · POSITION AND EVIDENCE

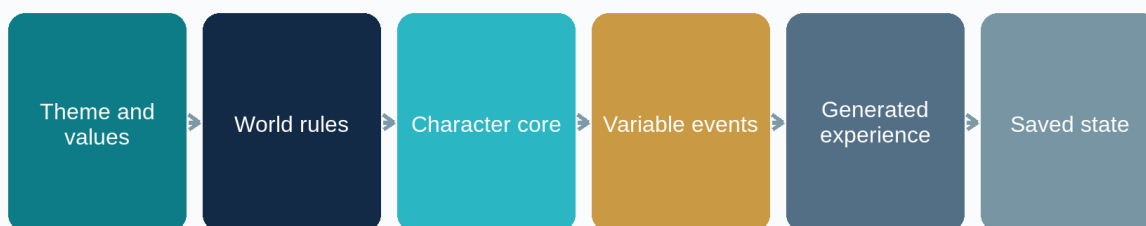
CHAPTER THESIS AI-native authorship can reside in rules, selection, orchestration, world design and responsibility. Instead of writing only a fixed script, a creator may construct a story constitution: themes, values, world laws, character cores, variable events, prohibited transformations and conditions for ending. Technical complexity does not itself create artistic significance.

INDUSTRY IMPLICATION Artistic evaluation should ask what problem the work makes perceptible, how generation changes form, whether variation is meaningful and whether the creator accepts responsibility for the system's consequences.

ARTISTIC JUDGMENT The author's contribution may be a space of possibility, but that space still requires intention, form and accountable choice.

From Story Constitution to Each Experience

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 16 · From Story Constitution to Each Experience

16.1 From script to story constitution

A story constitution preserves what must remain true while allowing the experience to vary. It defines theme, canon, character commitments, causal boundaries, sensitive content rules and the range of valid endings. Generation occurs inside this authored space. The constitution does not replace scenes or dialogue; it provides a higher-order structure that keeps thousands of possible scenes recognisably part of the same work.

16.2 New artistic units

The unit of composition may become a rule, a relationship, an encounter, a memory or a designed transition between world states. Rhythm can emerge across sessions rather than minutes. A character performance may be evaluated through consistency under unplanned input. Critics and juries therefore need methods that consider the realised experience and the system of possibility behind it.

16.3 Preserving works that cannot be exactly replayed

Preservation should capture the released model and version, asset set, story constitution, policy, important states, representative sessions and documentation of the runtime environment. A recording alone shows one performance but not the work's possibility space. Preservation institutions will need emulation, controlled replay and event archives alongside conventional masters.

Dimension	Critical question
Idea	What irreplaceable question does the work pose?
Form	Does generation create a new grammar?
Rules	Is the space of variation meaningful?
Relationship	Does audience action change understanding?
Responsibility	Does the creator understand the system's consequences?

TABLE 16 · CHAPTER ANALYTICAL FRAMEWORK

PART 05 INDUSTRY RESTRUCTURING

The convergence of film, games, live media and social worlds

PART PROPOSITION Shared world assets can be rendered through multiple media, making world operations more important than the final-file format.

PART 05 / GLOBAL AI CINEMA INDUSTRY 2026

CHAPTER 17 Film–Game Convergence as Shared-World Production

CHAPTER 17 · POSITION AND EVIDENCE

CHAPTER THESIS The next stage of film–game convergence is not a larger number of cross-promotions or adaptations. It is the construction of a shared world layer from which linear film, interactive games, generative live media, spatial experiences and community creation can be rendered. Characters, rules, state and provenance become common infrastructure.

INDUSTRY IMPLICATION World assets must be designed for reuse without flattening the distinct grammar of each medium or allowing one product's decisions to corrupt the shared canon.

INDUSTRY CONCLUSION The strategic move is from making the same IP twice to building one governed world that multiple media can read.

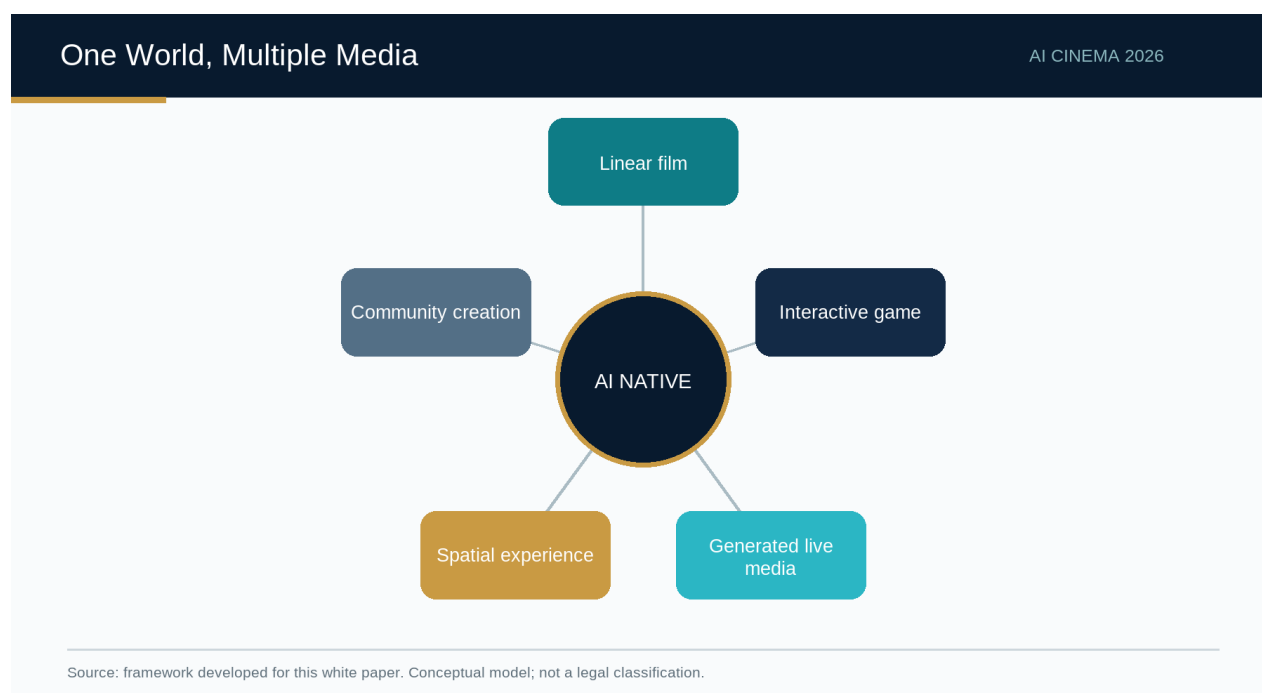


FIGURE 17 · One World, Multiple Media

17.1 From shared IP to shared state

Traditional adaptations share names, imagery and lore while rebuilding production assets and logic. A shared-world approach defines identity, geography, chronology, abilities, relationships and provenance in machine-readable layers. Film can interpret a character dramatically, a game can execute abilities, and live media can respond to events while remaining connected to a common version history. Reuse becomes semantic and operational, not merely visual.

17.2 The world studio

A world studio combines a story and art department with state architecture, asset systems, continuity engineering, live operations and governance. It distinguishes canonical state from temporary session state and community-created branches. The operating model resembles both a studio and a long-lived service

organisation: it releases works, maintains the world, resolves conflicts and plans how value returns to contributors.

17.3 Convergence risks

A shared world can become creatively rigid if every medium must obey a central database. It can also produce surveillance, exploitative labour, inaccessible canon or indefinite dependency on one platform. Governance should preserve medium-specific interpretation, forkable versions, contributor attribution and clear limits on what a commercial operator may change retroactively.

Shared layer	Film	Game / interaction	Governance
Character identity	Performance and arc	Capability and state	Consent / version
World semantics	Scene and theme	Map and rules	Origin / rights
Timeline	Narrative order	Quest and branch	Canon / fork
Behaviour	Performance choice	Executable action	Safety boundary

TABLE 17 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 18 Live Events, Competitions and Real-Time Cinema

CHAPTER 18 · POSITION AND EVIDENCE

CHAPTER THESIS Real-time AI cinema includes generated live shows, event-responsive worlds and persistent broadcast experiences. The core production challenge is not generating without interruption. It is maintaining control, intelligibility and recoverability under variable audience input, network conditions, model behaviour and safety events.

INDUSTRY IMPLICATION Live operations require rehearsed degradation, event logs and state recovery in the same way that conventional broadcast requires redundancy and editorial control.

OPERATING PRINCIPLE A real-time world is production-ready when it can fail safely and return to an explainable state.

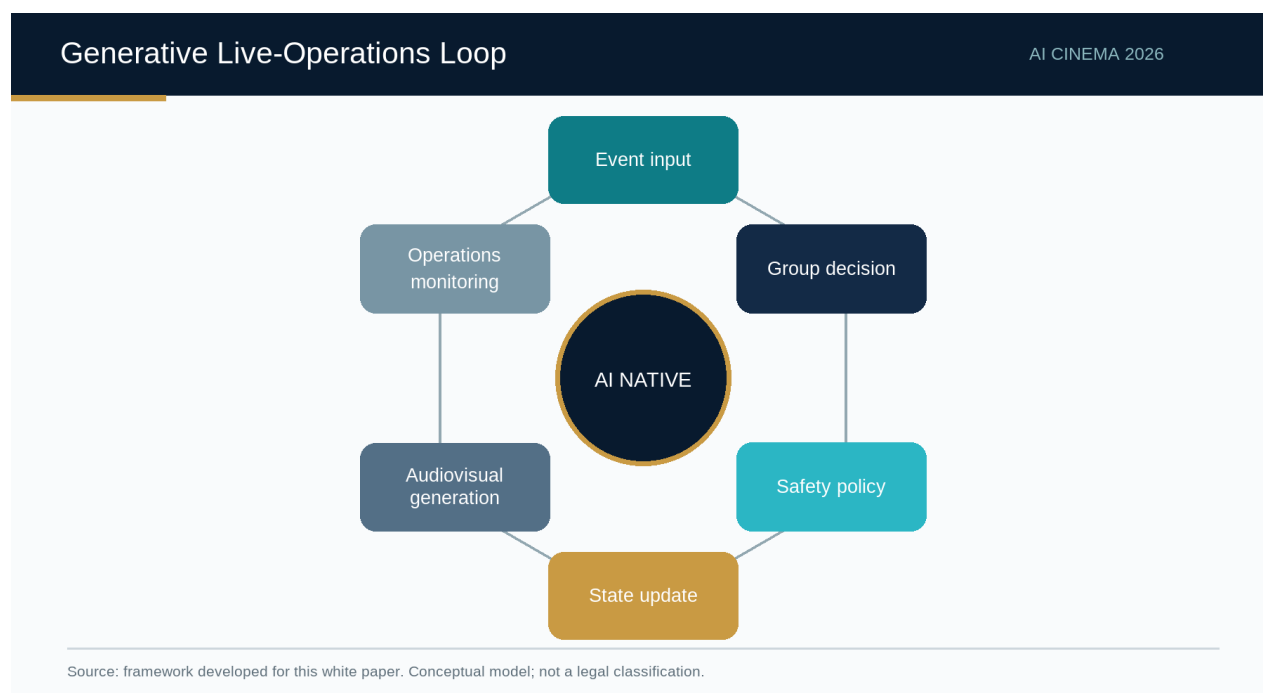


FIGURE 18 · Generative Live-Operations Loop

18.1 Three real-time product families

The first family is a scheduled generated programme with controlled participation. The second is an event layer that responds to sport, news, performance or community activity. The third is a persistent world that continues between major events. Each requires different latency, moderation and continuity policies. A scheduled show may use extensive pre-approved fallback; a persistent world must manage durable consequences and participant identity.

18.2 Audience input and governance latency

Safety review cannot always occur after generation because harmful output can be experienced immediately. The system should classify input, constrain the action space, select approved models and monitor the generated result. Governance latency is part of the end-to-end budget. When confidence is low, the world should reduce autonomy, delay publication or switch to a known-safe response.

18.3 Commercial and operational metrics

Key metrics include end-to-end latency, fallback rate, unexplained interruption, state-recovery time, incident frequency, cost per session minute, participant retention and world-state integrity. Growth can hide structural losses if every additional participant increases generation cost faster than revenue. Real-time products need service-level economics, not only audience metrics.

Metric	Purpose	Red line
End-to-end latency	Natural interaction	Persistent breach
Fallback rate	Service stability	No safe fallback
Safety interruption	Risk monitoring	Unexplained interruption
State recovery time	Incident handling	No rollback
Cost per minute	Unit economics	Loss grows with success

TABLE 18 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 19 A New Value Chain and New Professions

CHAPTER 19 · POSITION AND EVIDENCE

CHAPTER THESIS The AI cinema value chain adds layers for models, data, compute, world and workflow systems, content and production, distribution and devices, transactions and settlement, and continuous operations and safety. New work appears at the interfaces. The scarce professional is not the person who knows one model best, but the person who can integrate art, systems, assets and responsibility.

INDUSTRY IMPLICATION Training should move from tool demonstrations toward durable cross-domain capability: creative direction, system design, evidence, rights, evaluation and operational judgment.

TALENT JUDGMENT Model literacy is temporary advantage; orchestration, accountability and world-building are durable professional capabilities.

The New AI Cinema Value Chain

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 19 · The New AI Cinema Value Chain

19.1 Six layers of the emerging value chain

Foundation infrastructure provides models, data and compute. World and workflow systems organise state and tools. Production creates works and reusable assets. Distribution connects audiences, devices and contexts. Transaction systems manage permission and value flow. Operations maintain identity, safety, continuity and community. Organisations may occupy several layers, but conflicts of interest and responsibility should remain visible.

19.2 New professional combinations

An AI-native producer combines creative, budget and system responsibility. A world architect models rules and state. A continuity engineer manages identity and causality across generations. A provenance engineer designs records and metadata. A generative safety director defines refusal, degradation and recovery. These roles are not fixed job titles; they identify capability gaps that conventional departmental boundaries often leave unowned.

19.3 Labour transition and fair distribution

Automation can remove repetitive work while concentrating leverage in models and platforms. Transition programmes should track income, bargaining power, credit, reskilling and access to tools—not only the percentage of automated tasks. Contributors need clear attribution and remuneration when assets, performances and workflows continue to create value across products and years.

Role	Core responsibility	Capability
AI-native producer	Human–AI flow and budget	Creative + systems
World architect	Rules and state	Narrative + data modelling
Continuity engineer	Identity and causality	Quality + versioning
Provenance engineer	Evidence and fields	Standards + rights
Generative safety director	Refusal and recovery	Content + risk

TABLE 19 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 20 Market and Business-Model Transformation

CHAPTER 20 · POSITION AND EVIDENCE

CHAPTER THESIS Revenue can expand from tickets, advertising and licences toward world subscriptions, character subscriptions, interactive events, enterprise environments, asset services and agent-mediated requests. At the same time, continuous generation introduces cloud-style operating costs. A viable business must align narrative value, retention and permission with predictable unit economics.

INDUSTRY IMPLICATION Every persistent experience should know its cost per active user, per session minute and per accepted generated minute, and should separate growth in engagement from growth in unbounded compute demand.

FINANCIAL JUDGMENT Continuous media turns part of a studio into a service operator. World economics fail when operating cost scales faster than relationship value.

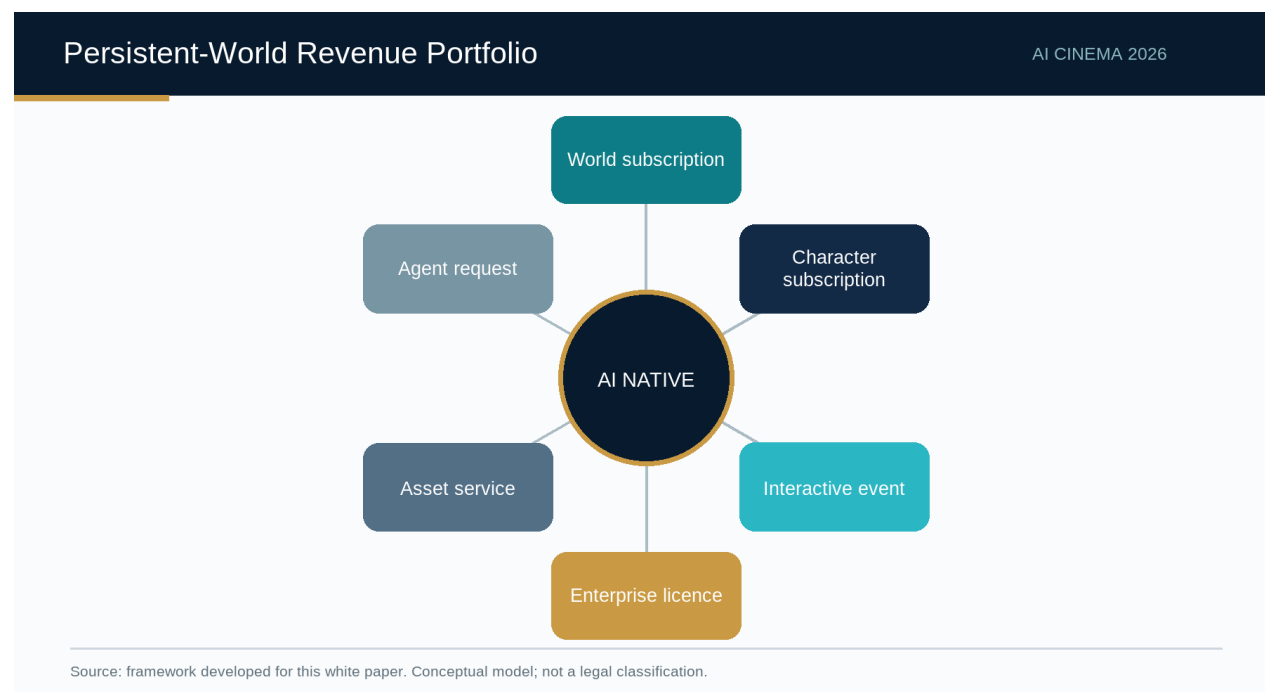


FIGURE 20 · Persistent-World Revenue Portfolio

20.1 Expanding the revenue unit

A world subscription sells maintained access and state. A character subscription sells continuity of relationship. An interactive event sells a time-bound experience. Enterprise licensing sells controlled brand or training environments. Asset services sell identity, workflow or world components. Agent requests support machine-initiated licensing or generation. Each unit requires different consent, service levels, accounting and refund logic.

20.2 From production budget to service unit economics

Traditional finance concentrates cost before release. Persistent worlds combine production expenditure with inference, storage, moderation, safety, customer support and state migration. The financial model should classify fixed world investment, variable session cost, acquisition cost, retention, liability reserves and end-of-life obligations. A product is not sustainable if success increases losses or makes withdrawal technically impossible.

20.3 Pricing and value distribution

Pricing should correspond to legible value rather than inferred emotional vulnerability. Revenue-sharing rules should address the continued use of performers, character creators, world designers, community contributors and workflow developers. Machine-readable contracts can improve settlement, but exceptions and disputes still need human process. The system should show which participant receives value when an asset is reused.

Model	Value unit	Required condition	Main risk
World subscription	Persistent state	Updates / retention	High operating cost
Character subscription	Long relationship	Consent / memory control	Dependence / manipulation
Interactive event	One experience	Reliable real-time	Peak capacity
Enterprise licence	Brand / environment	Control and provenance	Identity drift
Agent request	Machine transaction	Rights fields / payment	Agency liability

TABLE 20 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 21 The Emerging Global Industry Landscape

CHAPTER 21 · POSITION AND EVIDENCE

CHAPTER THESIS AI cinema is likely to develop through a multi-polar network rather than one global centre. North America has strength in foundation models, cloud and distribution; China and East Asia in application scale and supply chains; Europe in cultural institutions and rule-making; Japan and Korea in character, animation and games; and other regions in capital, youth markets and distinctive creative cultures.

INDUSTRY IMPLICATION A durable hub connects technology, assets, rules, talent and markets through trusted interfaces. It does not need to own every layer.

LANDSCAPE JUDGMENT The leading nodes will be those that make cross-border collaboration more trustworthy, portable and economically legible.

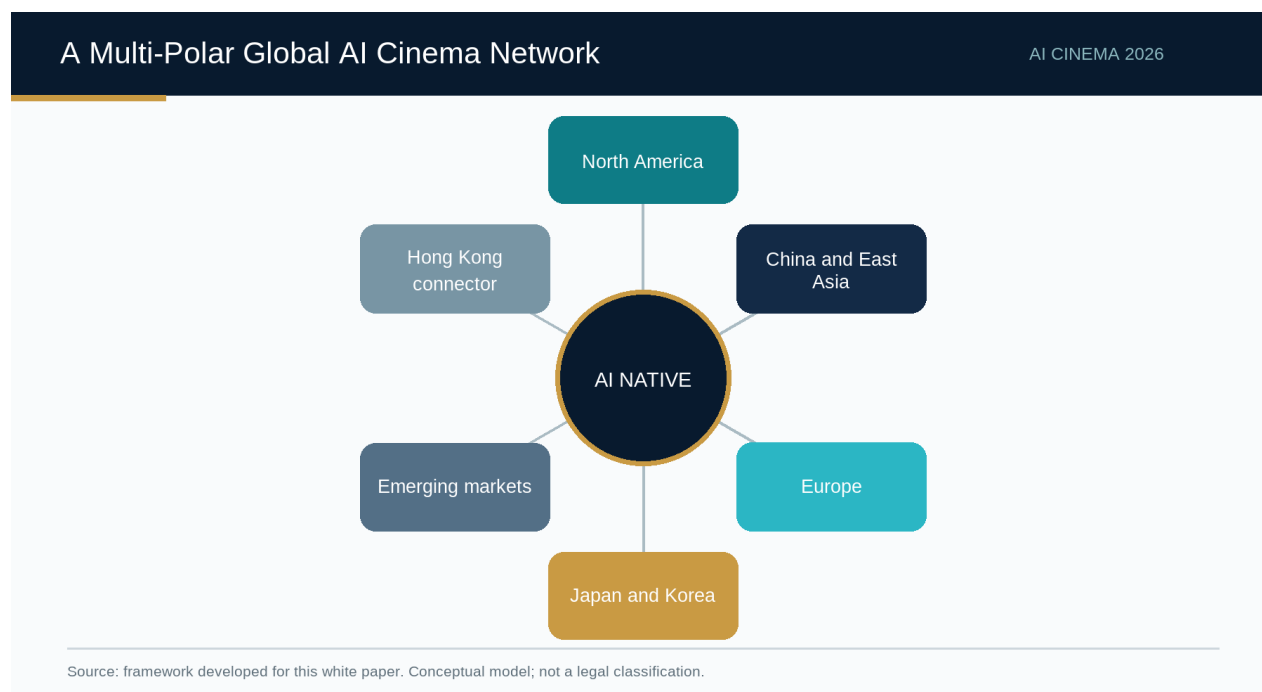


FIGURE 21 · A Multi-Polar Global AI Cinema Network

21.1 Competition is shaped by different endowments

Model capability is only one advantage. Access to production talent, local stories, game and animation expertise, data governance, capital, distribution, language markets and device supply chains all influence the field. A region can lead in world assets or governance even without training the largest model. The market will reward institutions that translate between these endowments.

21.2 Hong Kong's structural position

Hong Kong can contribute as a multilingual, cross-border market and standards connector with deep film culture, financial services and international networks. Its opportunity is not to imitate a hyperscale model centre. It is to make assets, projects, investment, festivals, evaluation and distribution interoperable across jurisdictions, while preserving a credible distinction between industry infrastructure and promotional claims.

21.3 A minimum language for international collaboration

Cross-border work needs shared definitions of work type, AI role, runtime behaviour, asset identity, permission scope, provenance status and responsibility. These fields should support local legal differences rather than pretending to replace them. A minimum common language allows creators and institutions to compare projects without forcing one jurisdiction's legal categories onto all participants.

Region	Strength	Opportunity	Constraint
North America	Models / cloud / distribution	Platforms and tools	Concentration / rights
China and East Asia	Applications / supply chain	Scaled content	Cross-border interoperability
Europe	Rules / culture	Trusted content	Capital / scale
Japan and Korea	Characters / animation / games	World assets	Language market
Hong Kong	Cross-border / finance / languages	Market and standards connection	Compute / production scale

TABLE 21 · CHAPTER ANALYTICAL FRAMEWORK

PART 06 DIGITAL ASSETS AND THE HUMAN-AI ECONOMY

From content delivery to machine-collaborative asset networks

PART PROPOSITION Value resides not only in a finished work, but also in workflows, characters, world states, behavioural rules and verifiable records.

PART 06 / GLOBAL AI CINEMA INDUSTRY 2026

CHAPTER 22 AI Audiovisual Assets: From a Finished Work to an Asset Family

CHAPTER 22 · POSITION AND EVIDENCE

CHAPTER THESIS The AI audiovisual asset family contains six distinct types: work, workflow, character, world, behaviour and record. Each needs a minimum identity, a version history, a value metric and a permission model. Treating all six as a single content file creates ambiguity and makes lawful reuse, evaluation and machine collaboration harder.

INDUSTRY IMPLICATION Indexing can identify and connect assets. It does not create, transfer or confirm legal rights, and it does not endorse an asset's quality or truth.

BOUNDARY STATEMENT An asset index describes infrastructure and relationships; legal title remains a matter for authorised records, evidence and applicable law.

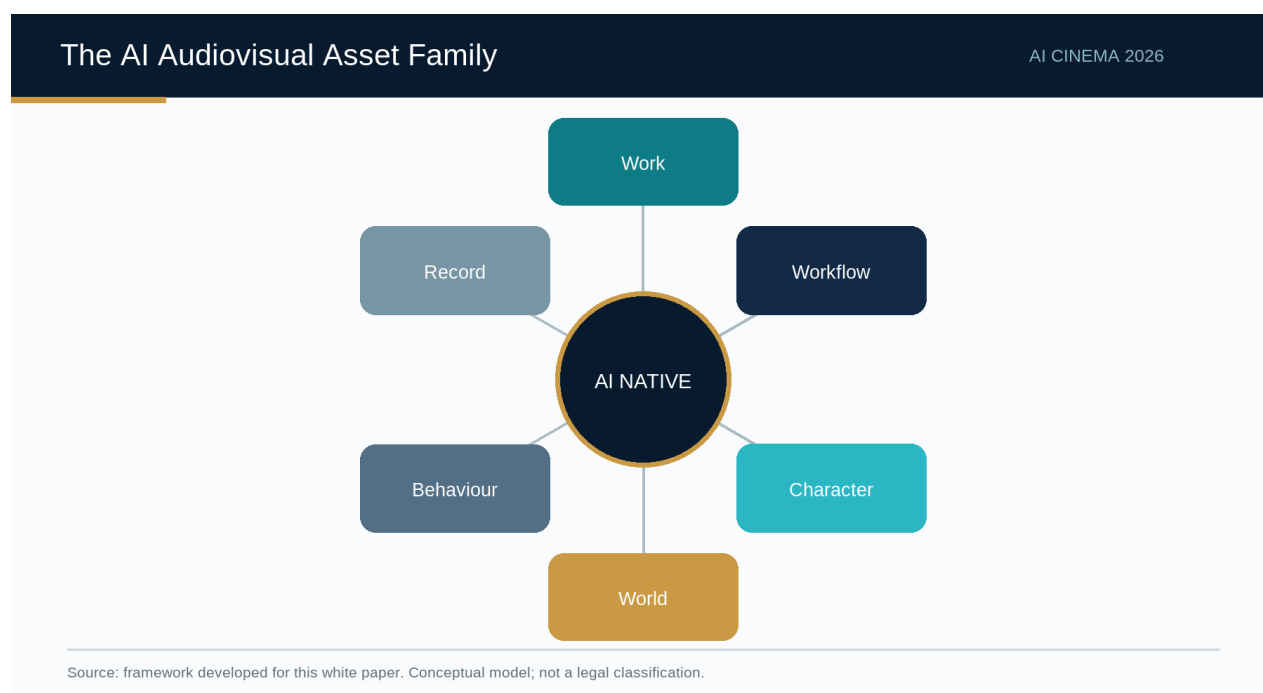


FIGURE 22 · The AI Audiovisual Asset Family

22.1 Six asset classes

A work is the released audiovisual experience. A workflow records tools, nodes and decisions. A character combines identity, model, performance and permission. A world contains semantics, rules and state. A behaviour defines an executable capability or interaction. A record provides provenance, event and lifecycle evidence. The same project may contain hundreds of related assets, each with different access and commercial terms.

22.2 Three conditions for usable assets

An asset must be identifiable: users and machines can distinguish it and its version. It must be interpretable: key technical, semantic and governance fields are readable. It must be usable within a known permission envelope. A visually impressive asset that lacks provenance, version clarity or permission scope may be valuable as inspiration but unusable in professional collaboration.

22.3 Evaluating asset value

Value may depend on audience performance, reliability, editability, relationship strength, reuse, provenance completeness and licence quality. One metric cannot compare all asset classes. A character can create durable value through identity stability; a workflow through acceptance rate; a world through extensibility; and a record through its ability to reduce transaction and dispute cost.

Asset	Minimum identity	Value metric	Permission focus
Work	Version / release	Audience / revenue	Exhibit / adapt
Workflow	Nodes / configuration	Acceptance / efficiency	Execute / redistribute
Character	Identity / model	Stability / relationship	Image / voice / behaviour
World	Rules / state	Extension / reuse	Enter / build / operate
Behaviour	Capability / interface	Control / safety	Invoke / delegate
Record	Provenance / event	Completeness / integrity	Inspect / retain

TABLE 22 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 23 The Lifecycle of AI Audiovisual Assets

CHAPTER 23 · POSITION AND EVIDENCE

CHAPTER THESIS A trusted asset lifecycle includes creation, declaration, identification, verification, evaluation, discovery, permission, use, settlement, change, archive and retirement. Trust does not mean that errors never occur. It means that changes, disputes, withdrawal and correction can be seen and handled without destroying the historical record.

INDUSTRY IMPLICATION Lifecycle design should separate public metadata, restricted operational information and private evidence, and should preserve event history when an asset changes status.

INSTITUTIONAL PRINCIPLE A credible registry records correction and withdrawal as first-class states rather than erasing inconvenient history.

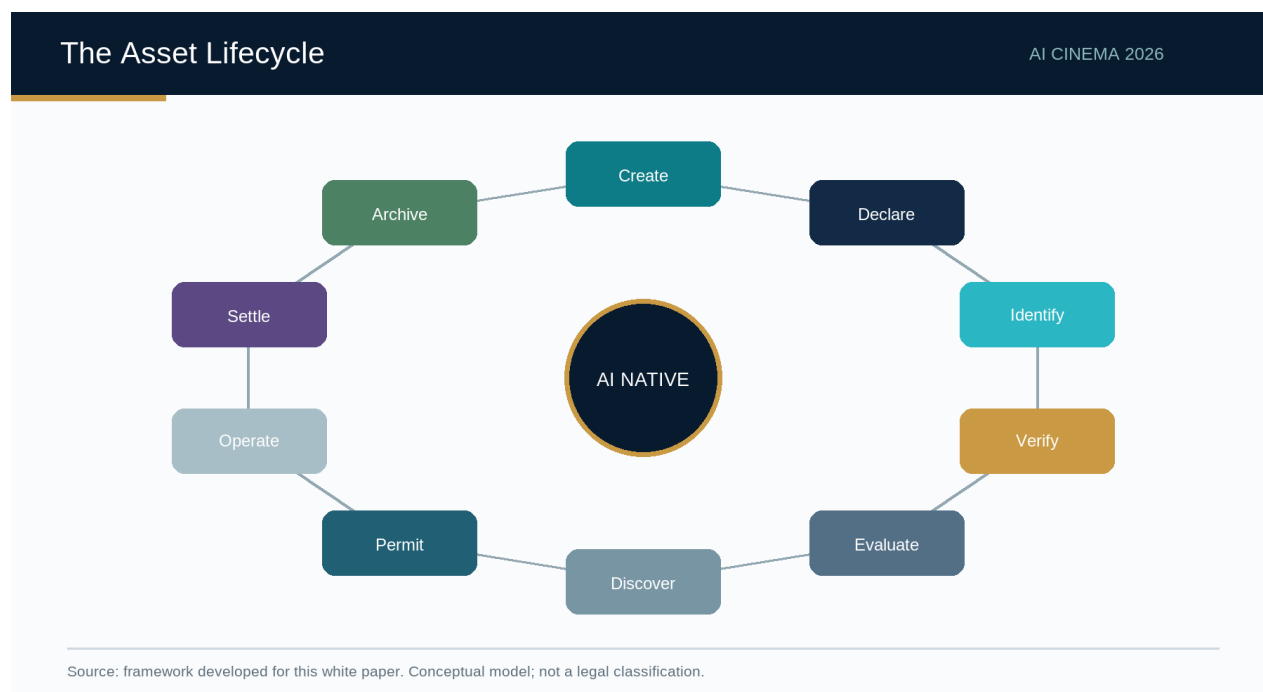


FIGURE 23 · The Asset Lifecycle

23.1 From creation to retirement

Creation produces an asset and initial evidence. Declaration identifies the responsible submitter. Identification assigns a persistent ID and parent relationships. Verification checks statements or signatures. Evaluation adds independent observations. Discovery exposes permitted metadata. Permission enables use; operation creates events; settlement records value flow. Change, archive and retirement preserve history and determine what remains accessible.

23.2 The minimum verifiable record

The record should include asset ID, type, version, parent asset, responsible contact, provenance summary, model and tool declaration, rights summary, governance status, technical format, integrity reference and dispute or withdrawal state. Not every field should be public. The minimum public layer should allow identification and informed contact without exposing confidential prompts, personal information or security-sensitive logs.

23.3 Withdrawal, disputes and orphan assets

A rights holder may withdraw permission; a submitter may disappear; an asset may be disputed or technically obsolete. The system needs states for suspension, contested status, replacement, restricted access and orphaned responsibility. Historical records should not imply current permission. Machine clients must be able to recognise that an asset once existed but cannot presently be used.

Field group	Example	Public layer
Identity	ID / type / version / parent	Public
Responsible parties	Creator / submitter / contact	Tiered
Provenance	Model / tool / asset declaration	Summary
Rights	Territory / term / use / status	Summary
Governance	Label / dispute / withdrawal	Public
Technical	Format / interface / hash	On request

TABLE 23 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 24 Agent-Readable Assets and Machine-Mediated Transactions

CHAPTER 24 · POSITION AND EVIDENCE

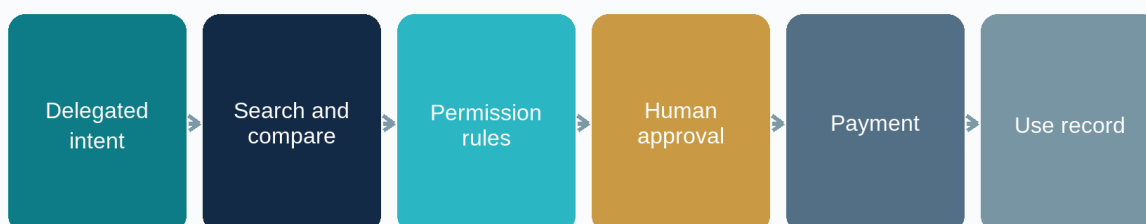
CHAPTER THESIS Agents can search for assets, compare terms, request permission and initiate payment. For this market to be safe, a probabilistic interpretation of a user's intent must remain separate from deterministic transaction authority. High-value, irreversible or personality-related rights require explicit approval by an identifiable principal.

INDUSTRY IMPLICATION Machine-readable permission should accelerate ordinary transactions without turning ambiguity into automatic consent or making responsibility disappear into an agent chain.

TRANSACTION FLOOR Agents may recommend and negotiate; sensitive rights and settlement require deterministic controls and accountable authorisation.

Layers of Agent-Mediated Asset Transactions

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 24 · Layers of Agent-Mediated Asset Transactions

24.1 The machine-readable rights envelope

An asset can express permitted uses, excluded uses, territory, duration, identity scope, price logic, attribution, technical constraints and whether human approval is required. These fields support discovery and preliminary negotiation. They do not remove exceptions, moral rights, collective agreements or jurisdictional rules. The envelope should state when machine interpretation must stop and a human decision must begin.

24.2 Separating probabilistic intent from deterministic settlement

A language model may infer that a user wants a licence, but inference is not authority. A transaction layer should present the exact asset, rights, price, counterparty and revocation conditions, obtain an approved authorisation and generate a durable record. Emerging agent-payment protocols are important infrastructure signals, while responsibility and consumer protection remain open design questions [R21–R22].

24.3 From a search market to a request market

Human users browse catalogues; agents can express structured requirements and ask the market to assemble a compliant solution. This can reduce transaction cost for localisation, asset replacement, music, voices and world components. It can also intensify concentration if one platform controls discovery, identity and payment. Open request formats and exportable records are therefore strategic infrastructure.

Field	Machine-readable value	Human handling
Use	View / generate / adapt	Unlisted use
Territory / term	Code / date	Conflict / exception
Identity	Character / performer scope	Sensitive person
Price	Fixed / tier / quote	High-value negotiation
Approval	Automatic / human	Personality / irreversible

TABLE 24 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 25 Connecting Human and AI Value: Dual-Domain Value Accounts

CHAPTER 25 · POSITION AND EVIDENCE

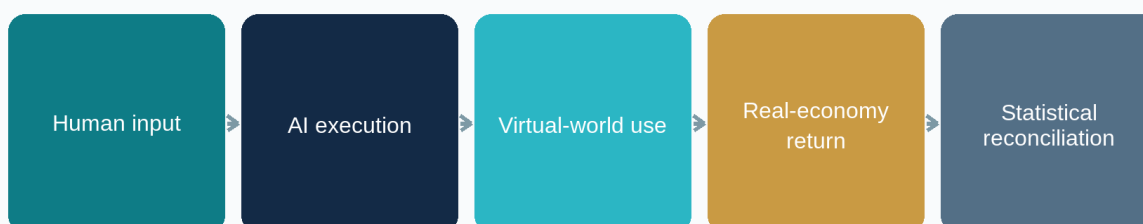
CHAPTER THESIS The future economy of AI worlds may be large, but transaction volume, asset valuation and economic value added are not interchangeable. Treating every internal agent action or virtual transfer as additional GDP would produce severe double counting. Measurement should identify human inputs, AI execution, virtual use, real-world return and intermediate consumption.

INDUSTRY IMPLICATION The proposed accounts are analytical tools for firms, investors and researchers. They are not an official second GDP and should remain compatible with national-accounting logic [R22–R23].

MEASUREMENT PRINCIPLE The larger the virtual economy becomes, the more disciplined its accounts must be.

Human–AI Dual-Domain Value Flow

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 25 · Human–AI Dual-Domain Value Flow

25.1 Why ‘AI-world GDP’ is not a direct additive total

An agent may call several models, purchase internal services and transfer a virtual asset before one final user pays for an experience. Counting every step as new value would count intermediate activity several times. Asset prices can also rise without corresponding current production. Measurement must distinguish gross transactions, revenue, income, capital formation, final use and value added, and must identify the economic principal behind automated activity.

25.2 Human–AI Dual-Domain Value Accounts

The human domain records creative labour, data preparation, capital, energy, rights and institutional services. The AI execution domain records inference and agent services attributed to a commissioning principal. The virtual-world domain records final use, maintained assets and internal circulation. The real-return domain records observable revenue, wages, licences and tax. Reconciliation removes intermediate transfers and shows how value and risk move between domains.

25.3 An AI-native audiovisual satellite account

A satellite account can estimate industry output, compensation, intermediate inputs, asset formation, imports, exports and energy use while preserving the main national accounts. Firms could begin with a voluntary sample: classify project expenditure, cloud operations, reusable world assets, human compensation and final revenue. The objective is a comparable evidence base, not a headline number detached from accounting boundaries.

Account	Records	How double counting is avoided
Human input	Creation / data / capital / energy	Mark intermediate input
AI execution	Inference / agent service	Attribute to principal
Virtual world	Final use / asset service	Separate internal circulation
Real return	Revenue / wage / licence / tax	Count observable flow
Satellite account	Value added / asset formation	Align to SNA

TABLE 25 · CHAPTER ANALYTICAL FRAMEWORK

PART 07 RIGHTS, GOVERNANCE AND GLOBAL STANDARDS

Transparency, responsibility and human agency

PART PROPOSITION Trust requires a traceable chain that connects provenance, permission, consent, version history and accountable release.

CHAPTER 26 Creative Control, Copyright and Digital Replicas

CHAPTER 26 · POSITION AND EVIDENCE

CHAPTER THESIS AI cinema implicates copyright, contract, personality rights, performer protections and platform terms. The relevant questions differ across training, project inputs, generated outputs, human selection and arrangement, and the release of an autonomous digital replica. Classification and provenance can supply facts, but legal effect depends on jurisdiction and evidence.

INDUSTRY IMPLICATION Rights review should be layered so that a lawful output claim is not used to conceal an unlawful input, an unauthorised replica or a contract restriction.

LEGAL BOUNDARY This paper provides industry frameworks, not legal determinations. Rights and remedies remain matters for applicable law and authorised decision-makers.

Four Rights Layers in AI Cinema

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 26 · Four Rights Layers in AI Cinema

26.1 Authorship: tool use and creative control

The U.S. Copyright Office and other authorities emphasise human authorship and control when assessing copyrightability [R15]. In AI-native work, evidence may include the creator's story constitution, selection, arrangement, revision and final approval. Merely requesting an unpredictable result is not equivalent to controlling expression. Different jurisdictions may reach different conclusions, so project records should support factual analysis without assuming a universal legal answer.

26.2 Separating training, project input and output

Training permission concerns the development of a model. Project-input permission concerns the assets supplied for a specific work. Output questions concern similarity, protected elements, contractual terms and human contribution. These layers should be reviewed independently [R16]. A production cannot cure an uncertain training or replica issue merely by placing a label on the final file.

26.3 Digital replicas and continuing consent

Consent for a face, voice or performance should specify media, context, territory, duration, autonomy, remuneration and withdrawal. A material change—such as allowing a replica to speak unscripted, enter a sensitive scene or operate after a performer is no longer available—should trigger renewed consent. The system must also maintain a practical means to stop new use and preserve the record of prior authorised releases.

Permission	Minimum content	Renewed-consent trigger
Data	Origin / purpose / retention	New training use
Face and voice	Media / context / territory	Sensitive scene
Autonomy	May say / do / prohibited	Passive to agent
Term and withdrawal	Stop / delete / archive	Model permanence
Compensation	Fixed / use / share	New revenue model

TABLE 26 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 27 Provenance, Content Labelling and the Responsibility Chain

CHAPTER 27 · POSITION AND EVIDENCE

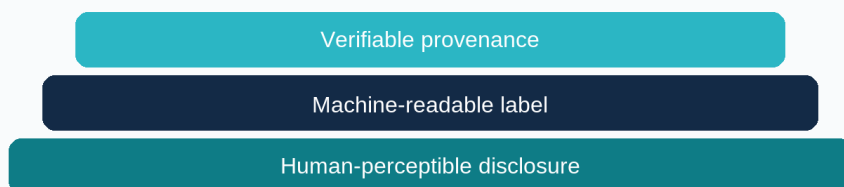
CHAPTER THESIS Trust requires three coordinated layers: verifiable provenance records, machine-readable labels and human-perceptible disclosure. These layers serve different purposes. Provenance can demonstrate association and integrity; a label can support machine and user awareness; neither automatically determines truth, legality or artistic value [R11].

INDUSTRY IMPLICATION Responsibility should be allocated across model provider, creator, producer, distributor and runtime operator according to control, rather than hidden behind a generic statement that AI was involved.

C2PA PRINCIPLE A provenance assertion can be verified as associated and untampered; the specification does not judge whether the content is true, good or lawful [R11].

Three Transparency Layers

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 27 · Three Transparency Layers

27.1 A three-layer transparency architecture

The provenance layer records assets, tools, responsible parties and integrity references. The machine-readable layer communicates generation and modification status to platforms and agents. The perceptible layer informs users in a context-appropriate form. China's labelling measures and EU transparency work both indicate movement toward coordinated explicit and implicit disclosure [R12–R14]. Implementation should remain accessible without overwhelming the audience.

27.2 A responsibility chain rather than a responsibility cloud

A model provider controls capability, safeguards and version disclosure. A creator controls inputs and creative choices. A producer controls workflow, rights and approval. A distributor controls labels, audience and platform policy. A runtime operator controls live policy, incident response and rollback. Several parties may share responsibility, but each should preserve the records that correspond to its control.

27.3 The institutional position of Index.Film™

Index.Film™ supports identification, recording, search and collaboration across work, workflow, character, world, behaviour and record assets. It should publish its metadata rules, correction process, access layers and dispute states. Inclusion is not endorsement; indexing does not confirm title; and an identifier does not replace legal registration or adjudication. These limitations are a source of credibility, not weakness.

Party	Primary control	Minimum record
Model provider	Capability / safety / version	Model card / change
Creator	Inputs / selection / revision	Creative evidence
Producer	Workflow / rights / approval	Delivery schedule
Distributor	Label / audience / platform	Release responsibility
Operator	Live policy / incident	Event and rollback

TABLE 27 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 28 Cultural Diversity, Human Agency and Sustainability

CHAPTER 28 · POSITION AND EVIDENCE

CHAPTER THESIS The success of AI cinema cannot be measured by the volume of generated content. It should be evaluated through cultural diversity, creator capability and livelihood, user agency, and the energy and material cost of production and operation. A more efficient model can still drive greater total consumption if usage expands rapidly [R19–R20].

INDUSTRY IMPLICATION Sustainability metrics should use the unit that matters—such as energy per accepted finished second or per session minute—while also reporting total demand and the distribution of benefit.

PUBLIC VALUE A productive AI cinema economy should increase expressive diversity and human capability without making resource use or platform dependence invisible.

Four Dimensions of Public Value

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 28 · Four Dimensions of Public Value

28.1 Cultural diversity and data concentration

Large models can reproduce dominant languages, aesthetics and commercial assumptions. Diversity requires more than translating an interface. Institutions should examine whose archives, performance traditions and narrative structures appear in data and evaluation; support local creators and languages; and enable communities to define sensitive uses. UNESCO's work on AI and culture underscores the need to connect innovation with cultural rights and plural participation [R18].

28.2 Human agency and the creator economy

Creators need meaningful control over tools, credit, income, portability and refusal. Audiences need the ability to understand, correct and exit adaptive systems. A platform can increase nominal participation while reducing bargaining power or making creators dependent on opaque ranking and generation services. Human agency is therefore an economic and institutional metric, not only an ethical aspiration.

28.3 Energy and sustainable generation

Evaluation should report energy per accepted output, per user session and per active world, while total energy and regional grid effects remain visible. Model efficiency, caching, edge placement, workload scheduling, renewable supply and selective generation can reduce demand. The industry should avoid the assumption that generated abundance has no physical infrastructure cost [R19–R20].

Dimension	Useful metric	Insufficient metric
Energy	Per accepted second / session minute	One inference
Culture	Language / region / creator distribution	Total volume
Labour	Income / transition / bargaining	Automation ratio
Agency	Control / exit / complaint	Time spent

TABLE 28 · CHAPTER ANALYTICAL FRAMEWORK

PART 08 FORECASTS, SCENARIOS AND ACTION

The 2026–2035 decision window

PART PROPOSITION A forecast is not a promise; it is a public account of assumptions, adoption gates, uncertainty and decision points.

CHAPTER 29 A 2026–2035 Technology and Industry Roadmap

CHAPTER 29 · POSITION AND EVIDENCE

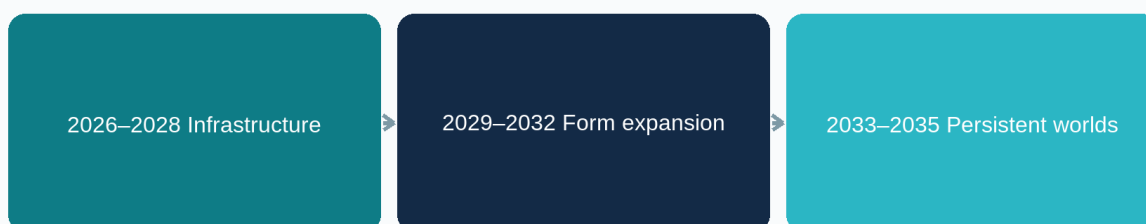
CHAPTER THESIS The roadmap uses observation windows rather than promised dates. The infrastructure period prioritises controllable generation, evidence and workflow integration. The form-expansion period depends on lower-cost real-time, multi-user and spatial capability. The persistent-world period requires cross-device continuity, stable economics and cross-border governance.

INDUSTRY IMPLICATION A form scales only when five gates align: technical reliability, economic viability, device access, governance legitimacy and cultural acceptance.

FORECAST DISCIPLINE Dates are review windows, not certainty. Capabilities that miss a gate should be delayed or narrowed rather than promoted as inevitable.

Three Observation Windows, 2026–2035

AI CINEMA 2026



Source: framework developed for this white paper. Conceptual model; not a legal classification.

FIGURE 29 · Three Observation Windows, 2026–2035

29.1 2026–2028: the infrastructure period

The priority is controllable generation, long-form state systems, native audio-video workflows, provenance, labelling and professional evaluation. AI-native films become more common, while real-time products remain bounded. Organisations that build portable assets, rights records, model-routing and acceptable-output metrics create advantages that survive rapid model turnover.

29.2 2029–2032: the form-expansion period

If latency, cost and device gates improve, interactive cinema, generated live events, shared film–game worlds and spatial experiences can move from demonstrations to repeatable products. New subscription and event revenue may emerge. The critical governance shift is from file-level records to event-level responsibility and from content permission to agent and world permission.

29.3 2033–2035: the persistent-world period

If identity, multi-user state, portability, economics and governance mature together, persistent cultural worlds may become a significant medium. They will not replace films or games. They will provide a layer in which works, characters and communities continue across devices and forms. Cross-border disputes, statistical treatment and long-term stewardship will become central institutions.

Window	Capability gate	Market gate	Governance gate
2026–2028	Controllable generation / workflow	Professional adoption	Labels / provenance
2029–2032	Multi-user / spatial / lower cost	Subscription / cross-media	Agent permission / event record
2033–2035	Persistent cross-device world	Stable world economics	Cross-border responsibility / statistics

TABLE 29 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 30 Four Future Scenarios

CHAPTER 30 · POSITION AND EVIDENCE

CHAPTER THESIS Scenario planning does not select the most attractive future. It tests whether present decisions survive different combinations of model capability, openness, trust, economics and regulation. The four scenarios are governed growth, platform enclosure, trust crisis and open worlds.

INDUSTRY IMPLICATION No-regret strategies include portable assets, visible responsibility, verifiable records, multi-platform operation, sustainable unit economics and human talent that can work across systems.

STRATEGY METHOD The best plan is not optimised for one forecast. It preserves choice across several plausible futures.



FIGURE 30 · Four Future Scenarios

30.1 Governed growth

Capability and revenue improve while standards, labelling and responsibility mature. Professional AI-native production expands, and audiences accept adaptive experiences because control is legible. The winning strategy is investment in trusted assets, evaluation and interoperable governance rather than maximal autonomy.

30.2 Platform enclosure

A small number of integrated platforms control models, identity, distribution, payment and world state. Creative tools are powerful but export is costly and terms can change. Organisations need multi-platform architecture, contractual exit rights, independent records and assets that can be migrated without losing identity or permission history.

30.3 Trust crisis

High-profile harm, rights disputes, synthetic deception or safety incidents reduce audience and institutional confidence. Regulation tightens and insurance costs rise. Provenance, explicit human responsibility, controlled product scope and rehearsed incident response become conditions of market access rather than optional compliance.

30.4 Open worlds

Open standards, agent transactions and portable identity allow creators and communities to operate across multiple worlds. Innovation and plural participation increase, but so do cross-system abuse, fraud and dispute complexity. Opportunity depends on strong identity permissions, contestable governance, interoperable evidence and fair settlement.

Scenario	Leading signal	No-regret strategy
Governed growth	Open standards / better economics	Trusted assets
Platform enclosure	High switching cost / concentration	Export and multi-platform
Trust crisis	Incidents / litigation / churn	Provenance and human responsibility
Open worlds	Interoperability / agent trade	Identity, permission and dispute handling

TABLE 30 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 31 Action Guide for Industry Stakeholders

CHAPTER 31 · POSITION AND EVIDENCE

CHAPTER THESIS Stakeholders should begin with work that preserves value across models and scenarios. A 90-day plan identifies assets, rights, costs and responsibility. A 12-month plan builds state, evidence and governance pipelines. A 36-month plan develops world operations, cross-media products, sustainable revenue and institutional partnerships.

INDUSTRY IMPLICATION Strategy should prioritise portability and accountable operating capability before committing to a closed technical stack or an untested future form.

EXECUTION PRINCIPLE Build the foundations that remain valuable when a model, platform or forecast changes.

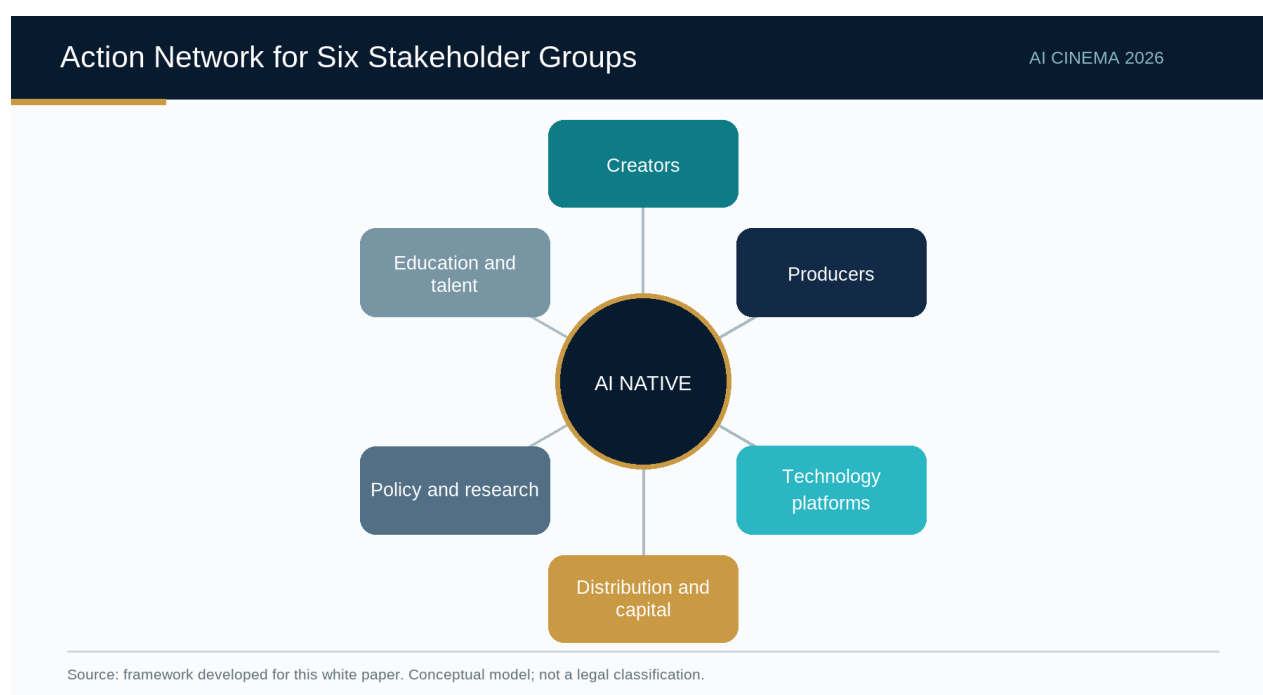


FIGURE 31 · Action Network for Six Stakeholder Groups

31.1 Creators and production organisations

Map story, character, world, workflow and evidence assets. Define AI-native creative intent before selecting tools. Establish state and version systems, acceptable-output metrics, rights review and final human approval. Run one bounded project that can be evaluated honestly rather than distributing experimentation across every production without learning.

31.2 Model, technology and platform providers

Disclose capability and material limitations; support export, version history and provenance; separate probabilistic proposals from transactions; provide safety degradation and incident tools; and publish changes that affect rights or output. Platforms should make responsible operation easier than opaque workarounds.

31.3 Distribution, investment and insurance

Use AI delivery schedules that identify assets, models, permissions, labels, evidence and runtime responsibility. Evaluate acceptable-output economics and persistent-world liabilities. Price risk according to

control and records rather than treating all AI involvement as identical. Require a practical exit and archive plan.

31.4 Policy, research and education

Develop shared terminology, representative test sets, public-interest research and statistical pilots. Education should combine cinema, interaction, systems, data, rights and operations. Certification should recognise demonstrated capability and professional responsibility rather than short-term familiarity with one interface.

Stakeholder	First 90 days	Within 12 months
Creators / producers	Asset and rights inventory	World-state and evidence pipeline
Technology platforms	Disclose capability and limits	Exportable provenance / fallback
Distribution / capital	AI delivery schedule	Unit economics / risk benchmark
Policy / research	Terms and samples	Public testing / statistical pilot
Education	Cross-domain curriculum	Role certification / transition

TABLE 31 · CHAPTER ANALYTICAL FRAMEWORK

CHAPTER 32 A Joint Action Agenda

CHAPTER 32 · POSITION AND EVIDENCE

CHAPTER THESIS The three publishing institutions propose a continuing public programme: maintain the dual-axis classification, develop evaluation benchmarks, evolve Index.Film™, connect cities and talent, test an audiovisual satellite account and publish an annual evidence review. The value of the white paper lies in whether its claims and frameworks can be tested, corrected and jointly built.

INDUSTRY IMPLICATION Each programme must publish its governance boundary: classification is not a legal ruling, evaluation does not replace criticism, indexing is not endorsement and a satellite account is not official GDP.

SHARED COMMITMENT A white paper should be an accountable public process, not a one-time assertion of authority.



FIGURE 32 · Joint Public-Interest Programme

32.1 Definitions and evaluation as public infrastructure

The P/R model, G0–G5 stages and E1–E4 evidence language should be revised through cases and peer review. Benchmarks should test long-form continuity, control, real-time recovery, provenance and human agency while leaving artistic quality open to criticism, juries and audiences.

32.2 Index.Film™ and market collaboration

The initial programme should publish minimum fields for six asset types, access levels, correction and dispute processes, and machine-readable permission boundaries. Market pilots can test how identifiers and records reduce transaction cost without claiming to determine ownership or artistic value.

32.3 A global city and talent network

Cities can host joint production, testing, education, festivals, residencies and market access. Participation should be based on concrete projects and reciprocal benefit rather than nominal membership. The network should make local creative cultures visible while improving cross-border interoperability.

32.4 Research and annual publication

Each annual edition should update technical evidence, regulatory developments, market data and scenario signals. It should report errors and changed judgments. A standing research process can invite institutions, creators, performers, engineers and public-interest groups to test the framework and propose revisions.

Public programme	Initial output	Governance boundary
Dual-axis classification	Annual specification / cases	Not a legal ruling
Evaluation benchmark	Long-form / real-time tests	Not artistic judgment
Index.Film™	Minimum fields for six assets	Inclusion is not endorsement
City network	Joint projects / talent flow	Not a nominal alliance
Satellite account	Enterprise sample pilot	Not official GDP
Annual white paper	Forecast review / revision	Public version history

TABLE 32 · CHAPTER ANALYTICAL FRAMEWORK

Appendices

The appendices convert the white paper's concepts into practical terms, evidence requirements, metadata, revision protocols and action tools.

Appendix A Core Glossary

Term	Definition
AI-native cinema	An audiovisual work in which AI has a material role in core creative dimensions and the work cannot be fully separated from its AI-enabled process.
Creative Agency axis (P0–P4)	A classification of the substantive role of AI in narrative, image, sound, character and interaction.
Runtime Generation axis (R0–R4)	A classification of how much of the experienced work is generated during use.
51%+ marker	A qualification threshold for substantive creative leadership on the P axis; not a quality or legal score.
World model	A system that maintains and predicts action-dependent state, memory and causal relationships.
Story constitution	The themes, rules, character commitments, variation boundaries and governance that define a generative narrative space.
World state	Authoritative information about entities, locations, relationships, events and permissions at a given time.
Asset family	The connected set of work, workflow, character, world, behaviour and record assets.
Provenance	Information about origin, transformation, responsible parties and integrity associated with an asset.
Rights envelope	Machine-readable fields that describe permitted use and when human approval is required.
Human–AI Dual-Domain Value Accounts	An analytical framework separating human inputs, AI execution, virtual use and observable real-world returns.
AI-native audiovisual satellite account	A statistical framework for industry output and assets designed to connect with national-accounting principles.

Appendix B P/R Classification and Creative Evidence

P should be assessed across narrative, image, sound, character and interaction, using weights declared before review. The 51%+ marker determines P3 qualification only. R is assessed independently according to runtime generation.

Evidence group	Description	Example
Creative intent	Concept, theme, rules and approved versions	Proposal / story constitution / approval
Tool chain	Models, versions, nodes and settings	Workflow export / model list
Generation process	Prompts, references, candidates and rejection	Logs / timestamps / samples
Human control	Selection, arrangement, revision and approval	Edit history / decision record
Runtime	Input, state, policy and generation level	Session record / architecture

Appendix C Index.Film™ Minimum Metadata Profile

This profile supports identification, recording, search and collaboration. It does not confirm rights or endorse an asset. Public, restricted and private layers should remain separate.

Field group	Example	Public layer
Identity	ID / type / version / parent	Public
Responsible parties	Creator / submitter / contact	Tiered
Provenance	Model / tool / asset declaration	Summary
Rights	Territory / term / use / status	Summary
Governance	Label / dispute / withdrawal	Public
Technical	Format / interface / hash	On request

Appendix D Evidence, Forecast and Annual Revision Protocol

- Grade key claims E1–E4 and preserve an accessible primary source or evidence note.
- State technical, economic, device, governance and cultural assumptions for forecasts.
- Report confirmation, delay, downgrade, withdrawal and error in the following edition.
- Separate vendor-reported performance from independent validation.
- Preserve public version history for classifications, metadata and evaluation frameworks.

Appendix E 90-Day, 12-Month and 36-Month Action Framework

Horizon	Organisational capability	Product and market	Governance and evidence
90 days	Inventory assets, tools, rights and costs	Select one bounded, evaluable pilot	Assign responsibility and delivery evidence
12 months	Build state, routing and evidence systems	Launch a cross-media minimum product	Implement labelling, incidents and fallback
36 months	Develop world operations and agent capability	Establish sustainable revenue and city collaboration	Pilot event-level records and measurement

Appendix F Primary References and Sources

Primary research papers, standards bodies, regulators, international organisations and official model-developer publications are prioritised. Information was checked through 16 July 2026.

[R01] OpenAI. Video generation models as world simulators. 2024.

[R02] Google DeepMind. Veo 3.1 - model overview and capabilities. 2025.

[R03] Meta AI. Movie Gen: A Cast of Media Foundation Models. 2024.

- [R04] Tencent Hunyuan. HunyuanVideo: A Systematic Framework for Large Video Generative Models. 2024.
- [R05] Wan Team. Wan: Open and Advanced Large-Scale Video Generative Models. 2025.
- [R06] Zhang et al.. Generative AI for Film Creation: A Survey of Recent Advances. 2025.
- [R07] Google DeepMind. Genie 3 - real-time interactive world model. 2025/2026.
- [R08] Liu et al.. Towards Interactive Video World Modeling: Frontiers, Challenges, Benchmarks, and Future Trends. 2026.
- [R09] Wang et al.. Matrix-Game 3.0: Real-Time and Streaming Interactive World Model with Long-Horizon Memory. 2026.
- [R10] Wu et al.. Infinite-World: Scaling Interactive World Models to 1000-Frame Horizons. 2026.
- [R11] C2PA. Content Credentials Technical Specification 2.2 and Explainer. 2025/2026.
- [R12] Cyberspace Administration of China et al.. Measures for Labelling AI-Generated and Synthetic Content. 2025.
- [R13] European Commission. Code of Practice on Transparency of AI-Generated Content. 2026.
- [R14] European Union. Regulation (EU) 2024/1689, Article 50. 2024.
- [R15] U.S. Copyright Office. Copyright and Artificial Intelligence, Part 2: Copyrightability. 2025.
- [R16] WIPO. Generative AI: Navigating Intellectual Property. 2024.
- [R17] NIST. AI Risk Management Framework: Generative AI Profile (NIST AI 600-1). 2024/2026.
- [R18] UNESCO. Report of the Independent Expert Group on Artificial Intelligence and Culture. 2025.
- [R19] International Energy Agency. Energy and AI. 2025.
- [R20] International Energy Agency. Data centre electricity use surged in 2025. 2026.
- [R21] Google. Agent Payments Protocol (AP2). 2025.
- [R22] International Monetary Fund. How Agentic AI Will Reshape Payments. 2026.
- [R23] United Nations Statistics Division. System of National Accounts 2025. 2025.
- [R24] Google. Android XR glasses demonstration and platform overview. 2025.
- [R25] Meta. Introducing Orion, augmented reality glasses prototype. 2024/2025.
- [R26] Apple Developer. Foveated Streaming for visionOS. 2026.
- [R27] Khronos Group. OpenXR and Spatial Entities extensions. 2025.

Appendix G Publishing Institutions

International Artificial Intelligence Audiovisual Asset Indexing Standards Alliance (IAISA)

Audiovisual asset identity, machine-readable metadata, provenance and market collaboration interfaces.

International Artificial Intelligence Film Producers Association (FIAPFIA)

AI-native production systems, technology pathways, professional coordination, evaluation and producer networks.

Hong Kong Artificial Intelligence International Film Festival Association (HKAIFF Association)

AI-native film art, creator ecosystems, festival practice and international city collaboration.