RESEARCH PROPOSAL | AIGC FILMMAKING

IntentFlow

A Human-AI Workflow for AIGC Filmmaking

From script understanding to editable storyboard, video, and rough cut

Grounded in an independently developed workflow used in real AIGC production.

Applicant: Dai Xudong | Proposed supervisor: Prof. Anyi Rao | HKUST

ABSTRACT

Generative AI can produce scripts, images, music, and video clips, yet creative intent is often lost when creators move between disconnected tools and repeatedly translate decisions into prompts. This project investigates how AI can become an editable production partner rather than a sequence of isolated generators. Building on an implemented production-automation workflow independently designed for real AIGC projects, I propose IntentFlow: a human-in-the-loop agentic workflow centered on a shared narrative-constraint ledger. The existing workflow addresses repetitive prompt writing, cross-stage context loss, and manual handoff overhead; the proposed research will formalize these production problems and evaluate new coordination mechanisms. The ledger represents plot beats, character identity, spatial continuity, visual style, camera intent, timing, and sound cues, and propagates approved human decisions across script analysis, storyboard planning, asset generation, video generation, and rough-cut assembly. The system will retain provenance for each human edit, support local selective regeneration, and surface cross-shot conflicts before costly downstream generation. A mixed-method evaluation will compare prompt-only production, pipeline automation without a constraint ledger, and the full IntentFlow system. Outcomes will be assessed through completion time, iteration and regeneration behavior, expert-rated narrative and visual consistency, creativity support, workload, usability, perceived control, and creative ownership.

APPLICANT PREPARATION

Creative-domain expertise	Implemented system	Real-world testbed
Formal composition training; audiovisual production; music, editing, and narrative practice.	Self-designed multi-agent production workflow spanning scripts, storyboards, assets, video, and rough cut.	Two Creative AI ventures, a 25-person AIGC team, 50+ delivered projects, and an AI music product with about 1,600 DAU as of 1 Sep 2026.

PROBLEM STATEMENT

Current generative-film practice is usually coordinated through manual prompt copying and informal memory. When a director changes a character motivation, location, visual rule, or pacing decision, the change may not propagate consistently to storyboards, assets, shots, and edits. The research problem is therefore not only generation quality. It is the preservation, negotiation, and revision of creative intent across a multi-stage production process.

- Scope: short-form narrative sequences of approximately 60-90 seconds and 6-10 shots, enabling rigorous evaluation before later scaling.
- Positioning: an HCI and creative-support systems contribution, using existing generative models as components rather than claiming a new foundation model.

RELATED WORK AND RESEARCH GAP

Research stream	What prior work establishes	Remaining opportunity
Script support	ScriptViz retrieves reference imagery and supports fixed versus variable visual elements; its study with 15 scriptwriters demonstrates the value of controllable visual exploration [1].	Translate script-level decisions into persistent constraints that remain editable in downstream production.
Storyboard and pre-visualization	Virtual Dynamic Storyboard ranks camera and animation proposals [2]. CineVision supports director-cinematographer iteration and was evaluated with 24 participants [3].	Connect storyboard decisions to assets, generated shots, timing, and later human revisions.
Controllable generation	ControlNet and AnimateDiff improve conditional control and personalized motion generation [5,6].	Coordinate heterogeneous models while preserving narrative intent and provenance across stages.
Film-production workflows	A recent survey identifies consistency, controllability, fine-grained editing, and motion refinement as central creator needs [4].	Operationalize these needs as cross-stage interaction mechanisms and evaluate them in realistic creative tasks.

CORE GAP

Existing systems typically optimize a stage - writing, pre-visualization, or generation. The proposed research asks how an editable representation of intent and feedback can coordinate the whole decision chain while preserving human control.

RESEARCH QUESTIONS

RQ1	How can high-level narrative intent be represented as editable constraints that remain traceable from script to rough cut?
RQ2	Does a shared constraint ledger improve script-to-shot alignment and cross-shot consistency compared with prompt-only and unstructured pipeline workflows?
RQ3	How do stage-level human controls and feedback provenance affect workload, perceived control, creative ownership, trust, and satisfaction?
RQ4	Which collaboration patterns differ between novice and experienced creators, and what should an AI filmmaking agent automate, suggest, or leave to the human?

WORKING HYPOTHESES

- H1: Constraint-aware workflows will increase expert-rated narrative and visual consistency while reducing avoidable regeneration.
- H2: Selective regeneration with visible provenance will improve perceived control and creative ownership without increasing task time.
- H3: Experienced creators will use more localized overrides, whereas novices will rely more on system suggestions and constraint templates.

EXISTING RESEARCH FOUNDATION: SELF-DESIGNED PRODUCTION WORKFLOW

The applicant has independently designed and implemented a production-oriented automation workflow in response to recurring problems observed in real AIGC projects. It reduces repetitive prompt construction, preserves context across production stages, and supports targeted revision after human feedback. It is not presented as a finished research contribution; it is the technical testbed and field-grounded starting point for the proposed study. The implemented workflow includes the following pipeline and creator controls.

IMPLEMENTED CAPABILITIES

Layer	Observed implementation evidence	Research relevance
Agent orchestration	Coordinated ScriptAgent and ProductionAgent with tool calls and task-specific sub-processes.	Supports division of labor and staged creative collaboration.
Persistent memory	Short-term messages, summaries, and semantic retrieval through a local MiniLM ONNX model.	Provides a basis for cross-session and cross-stage continuity.
Externalized Skills	Core planning and production instructions stored as editable Markdown Skills, including genre-specific director guidance.	Makes creative policies inspectable and iteratively configurable.
Model-provider layer	Programmable adapters for text, image, video, and TTS services with project-level model mapping.	Enables model-agnostic comparison and practical deployment.
Creator intervention	Editable plans, storyboards, prompts, art style, assets, candidate shots, track structure, and video duration; selective regeneration is supported.	Creates observable human decision points for HCI study.
Production state	SQLite-backed projects, assets, generation tasks, and saved flow data.	Enables logging of iteration paths, cost, and decision provenance.

WHAT THE PROTOTYPE DOES NOT YET ESTABLISH

- It does not yet encode one shared, editable representation of narrative intent across all production stages.
- Production success and 50+ delivered company projects provide requirements and cases, but do not replace controlled research evaluation.
- There is not yet a formal benchmark, baseline comparison, ablation study, or peer-reviewed user study for the proposed mechanism.

FIELD-GROUNDED OBSERVATIONS TO TEST

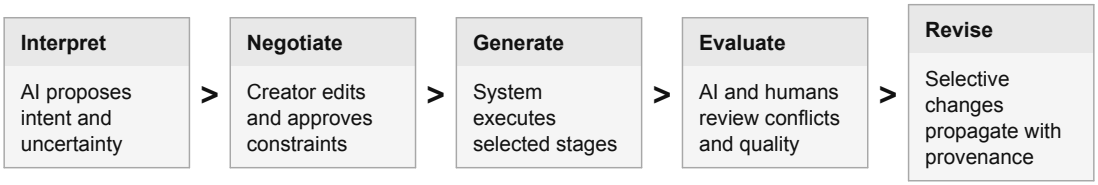
- Late-stage prompt repair is costly when upstream character, setting, or pacing decisions have drifted.
- Creators need to accept, reject, revise, pin, and selectively regenerate decisions instead of restarting an entire pipeline.
- Production teams require visible responsibility and decision history when multiple people and models contribute to a sequence.

PROPOSED SYSTEM: INTENTFLOW

IntentFlow extends the existing prototype with a shared representation and interaction layer. The goal is not full automation. The system should expose creative assumptions, make downstream consequences visible, and allow creators to intervene at the smallest useful scope.

Module	Function	Human control
1. Narrative Constraint Ledger	Represents plot beats, character identity, relationships, location, time, style, camera intent, pacing, and sound cues as typed constraints with confidence and scope.	Pin, edit, soften, override, or mark a constraint as exploratory.
2. Intent-to-Stage Compiler	Translates approved constraints into stage-specific plans and prompts for storyboards, assets, shots, sound, and rough cut.	Preview generated instructions before execution and choose which constraints propagate.
3. Consistency Monitor	Checks generated artifacts against active constraints using metadata, embeddings, and multimodal model judgments, with expert review retained for validity.	Inspect conflict explanations and accept, dismiss, or repair flagged inconsistencies.
4. Feedback Provenance Graph	Records who changed what, at which stage, and which downstream nodes depend on the decision.	Compare versions, restore earlier decisions, and selectively regenerate affected nodes.
5. Collaboration Policy	Controls whether the agent automates, recommends, asks, or waits based on uncertainty, cost, and creative importance.	Adjust autonomy and approval thresholds by project and stage.

TARGET INTERACTION LOOP



PROTOTYPE AND DATA SCOPE

- Unit of study: a 60-90 second fictional scene comprising 6-10 shots, rather than an unconstrained feature-length film.
- Content: licensed or participant-authored scripts; synthetic character and environment assets; no unauthorized personal likenesses.
- Models: model-agnostic interfaces to contemporary text, image, video, and audio services; model version, prompt, cost, latency, and output provenance will be logged.
- Research artifacts: anonymized interaction logs, constraint revisions, generated candidates, expert ratings, survey data, and interview transcripts, subject to licensing and ethics approval.

TECHNICAL DEVELOPMENT STRATEGY

- Use the existing TypeScript, Node.js, Electron, SQLite, Socket.IO, local ONNX retrieval, and multi-provider architecture as the engineering base.
- Add explicit schemas for narrative constraints and dependencies, versioned event logging, selective invalidation, and consistency reports.
- Separate model quality from interaction design through provider swaps, fixed seeds where available, recorded model versions, and ablation of each coordination mechanism.

MIXED-METHOD EVALUATION

Phase	Participants and task	Purpose
Formative pilot	6-8 film, animation, or AIGC creators complete one scene and join think-aloud interviews.	Refine constraint vocabulary, controls, failure explanations, and study procedure.
Controlled study	24-30 participants, balanced across students/novices and experienced creators; within-subject counterbalanced tasks.	Compare efficiency, consistency, usability, workload, control, ownership, and creativity support.
Field deployment	A smaller group of production practitioners uses the refined system on real short-form projects for 2-4 weeks.	Examine longitudinal adaptation, collaboration practices, and ecological validity.

EXPERIMENTAL CONDITIONS

A. Prompt-only	Separate generative tools with manual prompt transfer; represents common practice.
B. Pipeline automation	The applicant's existing production-automation workflow without the shared constraint ledger or provenance graph.
C. IntentFlow	Full constraint ledger, propagation, consistency monitoring, provenance, and selective regeneration.

MEASURES

Construct	Measures
Efficiency and behavior	Completion time, human actions, prompt revisions, regeneration count, accepted/rejected suggestions, stage revisits, inference cost, and latency.
Artifact quality	Expert-blind ratings of script-to-shot alignment, character and style consistency, spatial/temporal continuity, cinematic clarity, and overall narrative coherence.
Human experience	Creativity Support Index [8], NASA-TLX [9], usability/UEQ items, perceived control, trust calibration, creative ownership, and satisfaction.
Qualitative evidence	Think-aloud observations, semi-structured interviews, and thematic analysis of breakdowns, workarounds, and desired control boundaries.

ANALYSIS AND VALIDITY

- Use repeated-measures or mixed-effects models, report effect sizes and uncertainty, and account for participant experience, task, and model version.
- Use two or more independent expert raters; report inter-rater reliability and resolve rubric ambiguity through a pilot set.
- Conduct ablations for constraint propagation, consistency monitoring, and provenance to identify which mechanism produces each effect.
- Triangulate interaction logs, artifact ratings, questionnaires, and interviews; treat automated similarity scores as supporting measures rather than ground truth.

ETHICS, RIGHTS, AND RESPONSIBLE USE

- Obtain informed consent and institutional ethics approval before user studies; minimize collection of identifiable data.
- Use licensed, participant-authored, or appropriately permitted material; record model and asset provenance and respect withdrawal requests.
- Prohibit non-consensual likeness generation and misleading impersonation; include disclosure and audit trails for generated outputs.
- Study creative labor and authorship concerns directly, including when automation reduces agency or obscures human contribution.

FOUR-YEAR WORK PLAN

Period	Research activities	Planned outputs
Year 1	Systematic review; requirement interviews; stabilize prototype; define constraint schema; formative pilot; ethics submission.	Problem framing, design requirements, pilot dataset, prototype v1.
Year 2	Implement constraint ledger, provenance graph, and consistency monitor; conduct component benchmarks and ablations.	IntentFlow v2; technical/HCI paper submission; reproducible evaluation protocol.
Year 3	Controlled mixed-method study; refine interaction policies; deploy with production practitioners.	Main empirical paper; longitudinal findings; design guidelines.
Year 4	Synthesis, replication, dissertation writing, and selective release of code/data where rights permit.	Thesis, final publications, research toolkit and documented benchmark.

EXPECTED CONTRIBUTIONS

Technical and representational	HCI and creative practice
An editable narrative-constraint ledger, cross-stage dependency model, consistency monitor, and selective-regeneration mechanism for generative filmmaking.	Evidence about control, ownership, workload, trust, and collaboration patterns, resulting in design principles for creator-centered AI production systems.

ALIGNMENT WITH PROF. ANYI RAO AND HKUST AMC

The project directly connects Prof. Rao's research in AI for creativity, HCI, computer vision, graphics, film studies, script visualization, pre-visualization, controllable generation, and generative-film workflows [1-6]. It complements ScriptViz and CineVision by extending creator control from writing and pre-production into a traceable generation-to-editing loop. Within AMC, the work combines machine creativity, machine-assisted art practice, system building, and mixed-method evaluation. The applicant's composition background and deployed production systems contribute a practitioner-research perspective and a concrete environment for studying how AI becomes a creative collaborator.

SELECTED REFERENCES

[1] Rao, A., Chou, J.-P., and Agrawala, M. 2024. ScriptViz: A Visualization Tool to Aid Scriptwriting based on a Large Movie Database. UIST 2024, Article 21, 13 pages. DOI: 10.1145/3654777.3676402.

[2] Rao, A., Jiang, X., Guo, Y., Xu, L., Yang, L., Jin, L., Lin, D., and Dai, B. 2023. Dynamic Storyboard Generation in an Engine-based Virtual Environment for Video Production. ACM SIGGRAPH 2023 Posters. DOI: 10.1145/3588028.3603647.

[3] Wei, Z., Wu, H., Zhang, L., Xu, X., Zheng, Y., Hui, P., Agrawala, M., Qu, H., and Rao, A. 2025. CineVision: An Interactive Pre-visualization Storyboard System for Director-Cinematographer Collaboration. UIST 2025, Article 18, 18 pages. DOI: 10.1145/3746059.3747793.

[4] Zhang, R. et al. 2025. Generative AI for Film Creation: A Survey of Recent Advances. CVPR Workshops 2025, 6321-6333.

[5] Zhang, L., Rao, A., and Agrawala, M. 2023. Adding Conditional Control to Text-to-Image Diffusion Models. ICCV 2023, 3836-3847.

[6] Guo, Y., Yang, C., Rao, A., Liang, Z., Wang, Y., Qiao, Y., Agrawala, M., Lin, D., and Dai, B. 2024. AnimateDiff: Animate Your Personalized Text-to-Image Diffusion Models without Specific Tuning. ICLR 2024 (Spotlight).

[7] Amershi, S. et al. 2019. Guidelines for Human-AI Interaction. CHI 2019, Paper 3, 13 pages. DOI: 10.1145/3290605.3300233.

[8] Cherry, E. and Latulipe, C. 2014. Quantifying the Creativity Support of Digital Tools through the Creativity Support Index. ACM TOCHI 21(4), Article 21, 25 pages. DOI: 10.1145/2617588.

[9] Hart, S. G. and Staveland, L. E. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In Human Mental Workload, 139-183. DOI: 10.1016/S0166-4115(08)62386-9.