CINEAGI: CHARACTER-CONSISTENT MOVIE CREATION THROUGH LLM-ORCHESTRATED MULTI-MODAL GENERATION AND CROSS-SCENE INTEGRATION

TECHNICAL APPENDIX

Anonymous Submission

March 30, 2026

A Multi-Agent Coordination Algorithms

This section provides detailed algorithmic specifications for the multi-agent coordination framework introduced in Section 3.2 of the main paper. Our hierarchical agent system employs specialized Large Language Model (LLM) agents that coordinate through structured message passing to ensure narrative coherence and cross-modal consistency.

A.1 Agent Communication Protocol

We define a formal communication protocol $\mathcal{P}$ that governs interactions between agents in our narrative synthesis module. Each agent A_i maintains a state vector $\mathbf{s}_i(t) \in \mathbb{R}^d$ at time step t, and communicates through structured message passing to maintain consistency across character profiles, scene descriptions, and dialogue sequences.

Algorithm 1 Multi-Agent Coordination Protocol

Require: Initial story prompt $\mathcal{S}_0$, agent set $\mathcal{A} = \{A_1, A_2, ..., A_5\}$
Ensure: Validated narrative specification $\mathcal{N}^*$
Initialize agent states: $\mathbf{s}_i(0) \leftarrow \text{InitState}(A_i, \mathcal{S}_0)$
for $t = 1$ to T_{max} **do**
  for each agent $A_i \in \mathcal{A}$ **do**
    Receive messages: $\mathcal{M}_i(t) \leftarrow \text{Collect}(\{A_j : j \neq i\})$
    Update state: $\mathbf{s}_i(t) \leftarrow \text{Update}(\mathbf{s}_i(t-1), \mathcal{M}_i(t))$
    Generate output: $\mathcal{O}_i(t) \leftarrow \text{Generate}(A_i, \mathbf{s}_i(t))$
    Broadcast: $\text{Send}(\mathcal{O}_i(t), \mathcal{A} \setminus \{A_i\})$
  end for
  if $\text{QualityInspector}(\{\mathcal{O}_i(t)\}) = \text{True}$ **then**
    return $\mathcal{N}^* \leftarrow \text{Synthesize}(\{\mathcal{O}_i(t)\})$
  end if
end for

A.2 Coordination Effectiveness Analysis

Our multi-agent system achieves stable narrative synthesis through iterative refinement and quality validation. Rather than formal mathematical convergence, we analyze the coordination effectiveness through empirical validation of output consistency.

The coordination process follows a structured workflow where agents iteratively refine their outputs based on feedback from other agents. The Quality Inspector serves as the convergence criterion, evaluating whether the collective agent outputs meet predefined consistency thresholds across character profiles, scene descriptions, and dialogue coherence.

Empirically, we observe that the system typically reaches satisfactory narrative consistency within 3-7 communication rounds, as demonstrated in our hyperparameter sensitivity analysis (Table 5). The coordination effectiveness is measured through the improvement in Overall Consistency scores as agents exchange information and refine their outputs collaboratively.

A.3 Agent Specialization Details

Each agent in our system is designed with specific expertise and prompt templates to handle different aspects of movie creation while maintaining coordination with other agents.

Character Designer Prompt Template:

```
Act as a Character Designer for movie production. Your task is to analyze the input
    story and establish detailed character profiles with hierarchical attributes.

**Input**: Story concept text

**Output JSON Format**:
{
   "Character_profiles": {
      "Man1": {
         "appearance": {
            "facial_structure": "Detailed facial features description",
            "build": "Body type and physique",
            "distinguishing_features": "Unique physical characteristics",
            "clothing_style": "Typical attire and style preferences"
         },
         "personality": {
            "core_traits": ["trait1", "trait2", "trait3"],
            "behavioral_patterns": "How they typically act and react",
            "speech_patterns": "How they speak and express themselves",
            "emotional_tendencies": "Default emotional states and responses"
         },
         "background": {
            "occupation": "Professional background",
            "life_context": "Relevant personal history",
            "motivations": "What drives this character",
            "relationships": "Connections to other characters"
         }
      },
      "Woman1": {
         // Similar structure for each character
      }
   },
   "Character_relationships": {
      "dynamic_analysis": "How characters interact with each other",
      "conflict_points": "Areas of tension or disagreement",
      "emotional_bonds": "Positive connections and chemistry"
   }
}

**Requirements**:
1. Create comprehensive character specifications covering appearance, personality,
    and behavioral patterns
2. Ensure each character has unique and distinguishable traits
3. Establish clear character hierarchies and relationships
4. Provide consistent reference points for all downstream generation
5. Focus on visual and behavioral consistency across scenes
6. Include detection keywords for character identification

Please analyze the story and create detailed character profiles.
```

Script Writer Prompt Template:

```
Act as a Script Writer for movie production. Create shooting scripts centered around
    character profiles while maintaining strict alignment with established character
    descriptions.

**Input**:
- Character profiles from Character Designer
- Story concept text

**Output JSON Format**:
{
   "Scene_script": {
      "scene1": {
         "scene_text": "Detailed 5-second scene description using character
             appearance instead of identifiers. Focus on single-shot composition and
             character-centric description with technical details including camera
             angles, movements, and lighting",
         "technical_specs": {
            "camera_angle": "Description of camera positioning",
            "camera_movement": "Any camera motion during scene",
            "lighting": "Lighting setup and mood",
            "composition": "Visual composition details"
         },
         "seg_kws": {
            "Man1": "man",
            "Woman1": "woman"
         },
         "character_positioning": "Detailed description of where each character is
             positioned"
      },
      "scene2": {
         // Similar structure for each scene
      }
   },
   "Scene_transition": {
      "scene1_2": "Detailed visual description of transition including specific
          effects (dissolve, fade, wipe), lighting changes, camera movements, and
          technical feasibility",
      "scene2_3": "Next transition description"
   }
}

**Requirements**:
1. 20-second total duration split into 5-second scenes
2. Use character appearance descriptions, not identifiers
3. Maintain textual character consistency across scenes
4. Include technical cinematography specifications
5. Provide character detection keywords for downstream processing
6. Ensure scene descriptions align with character profiles
7. Create smooth, technically feasible transitions
8. Focus on single-shot composition per scene

Please create the shooting script based on the provided character profiles and story.
```

Storyteller Prompt Template:

```
Act as a Storyteller specializing in narrative flow and character dynamics. Create
    dialogue content with precise timing while maintaining character consistency and
    dramatic progression.

**Input**:
- Character profiles from Character Designer
- Scene scripts from Script Writer
- Story concept text

**Output JSON Format**:
```

```
{
   "Dialogue_script": {
      "scene1_dialogue": {
         "Man1": {
            "text": "Character dialogue matching their established speech patterns",
            "frame_range": [1, 30],
            "emotion": "Specific emotional state and speaking style",
            "character_alignment": "How this dialogue reflects character personality"
         },
         "Woman1": {
            "text": "Character dialogue content",
            "frame_range": [31, 60],
            "emotion": "Speaking style and emotional context",
            "character_alignment": "Personality consistency check"
         }
      },
      "scene2_dialogue": {
         // Similar structure for each scene
      }
   },
   "Narrative_flow": {
      "story_progression": "How the dialogue advances the plot",
      "character_development": "How characters evolve through dialogue",
      "emotional_arc": "Overall emotional journey of the story",
      "pacing_analysis": "Timing and rhythm of dialogue delivery"
   }
}

**Requirements**:
1. Ensure character actions align with established personalities
2. Maintain consistent character portrayal across scenes
3. Create natural dialogue for text-to-audio generation
4. Provide frame-level timing specifications (within 128 frames at 20fps)
5. Include realistic pacing and emotional progression
6. Preserve dramatic arcs and story coherence
7. Use innovative character dynamic analysis
8. Match dialogue to character's behavioral patterns and speech style

Please craft the narrative dialogue based on character profiles and scene
    descriptions.
```

Composer Prompt Template:

```
Act as a Composer and Music Director for movie production. Create cohesive musical
    direction that enhances audiovisual synchronization and emotional resonance across
    the entire production.

**Input**:
- Character profiles from Character Designer
- Scene scripts from Script Writer
- Dialogue scripts from Storyteller
- Story concept text

**Output JSON Format**:
{
   "BGM_script": {
      "scene1_bgm": {
         "music_description": "Detailed musical composition description",
         "technical_specs": {
            "tempo": "120 bpm",
            "key": "C major with modulation to G major",
            "style": "Cinematic orchestral with contemporary elements",
            "instruments": "String quartet, piano, subtle electronic ambient",
            "sound_characteristics": "Warm reverb, atmospheric, gentle crescendo",
            "intensity": "Medium emotional intensity building to high"
```

```
         },
         "emotional_context": "How music supports scene emotion",
         "character_themes": "Musical elements reflecting character personalities"
      },
      "scene1_2_transition_bgm": {
         "transition_music": "Musical bridge description",
         "technical_specs": {
            // Similar detailed specifications
         },
         "continuity": "How music maintains flow between scenes"
      }
   },
   "Musical_coherence": {
      "overall_theme": "Main musical motifs running through the production",
      "character_musical_identity": "Unique musical elements for each character",
      "emotional_progression": "How music evolves with story development",
      "synchronization_points": "Key moments requiring audio-visual alignment"
   }
}

**Requirements**:
1. Create cohesive musical direction enhancing audiovisual synchronization
2. Generate background music specifications rather than scene-by-scene generation
3. Ensure emotional resonance across entire production
4. Include specific technical details: tempo (bpm), style, instruments, key, sound
    characteristics
5. Maintain thematic coherence while adapting to scene dynamics
6. Consider character personalities in musical choices
7. Provide detailed specifications for music generation models
8. Create musical identity that supports narrative structure

Please compose the musical direction based on all provided narrative elements.
```

Quality Inspector Prompt Template:

```
Act as a Quality Inspector for movie production. Ensure end-to-end consistency by
    validating interconnections between all agent outputs and preventing cascading
    errors.

**Input**:
- Character profiles from Character Designer
- Scene scripts from Script Writer
- Dialogue scripts from Storyteller
- Musical direction from Composer
- Original story concept

**Output JSON Format**:
{
   "Quality_assessment": {
      "character_consistency": {
         "profile_alignment": "Check if characters match across all components",
         "appearance_consistency": "Verify visual descriptions are consistent",
         "behavioral_consistency": "Ensure personality traits are maintained",
         "issues_found": ["List any inconsistencies"],
         "corrections_needed": ["Specific fixes required"]
      },
      "narrative_coherence": {
         "story_alignment": "How well outputs match original concept",
         "scene_flow": "Logical progression between scenes",
         "dialogue_relevance": "Dialogue appropriateness to scenes",
         "issues_found": ["List any narrative problems"],
         "corrections_needed": ["Specific narrative fixes"]
      },
      "technical_validation": {
         "timing_consistency": "Frame ranges and duration alignment",
```

```
         "segmentation_keywords": "Validity of detection keywords",
         "transition_feasibility": "Technical viability of transitions",
         "music_synchronization": "Audio-visual timing compatibility",
         "issues_found": ["Technical problems identified"],
         "corrections_needed": ["Technical fixes required"]
      },
      "overall_coherence": {
         "cross_modal_alignment": "Integration between visual, audio, and narrative",
         "production_readiness": "Assessment of readiness for next stage",
         "quality_score": "Overall quality rating (1-10)",
         "recommendations": ["Suggestions for improvement"]
      }
   },
   "Standardized_output": {
      "validated_components": "Components that passed quality check",
      "corrected_elements": "Elements that were corrected",
      "final_json_structure": "Standardized format for downstream processing"
   }
}

**Requirements**:
1. Validate interconnections between all agent outputs
2. Prevent cascading errors from individual agents
3. Verify strong relevance to original user input
4. Examine coherence between character profiles, dialogue, scenes, and music
5. Check technical specifications for downstream processing
6. Ensure standardized JSON format for subsequent stages
7. Provide specific correction recommendations
8. Maintain critical oversight role in the pipeline

Please perform comprehensive quality inspection of all narrative synthesis components.
```

B Identity Preservation Mechanism

This section details the textual consistency mechanisms used to maintain character identity throughout the movie generation process, focusing on ensuring alignment between character portrait prompts and scene description prompts.

B.1 Character Description Consistency

Our identity preservation mechanism operates at the textual prompt level, ensuring that character descriptions remain consistent across different generation stages. The key challenge is maintaining coherence between the detailed character portraits generated by the Character Designer and the character descriptions embedded within scene scripts produced by the Script Writer.

Character Portrait Prompt Standardization: The Character Designer generates comprehensive character descriptions following a standardized template that captures essential visual and behavioral attributes. These descriptions serve as the canonical reference for character identity throughout the pipeline.

Scene-Character Alignment: The Script Writer ensures that character descriptions within scene scripts maintain consistency with the canonical character portraits. This is achieved through explicit cross-referencing and validation mechanisms that compare character attributes across different textual contexts.

B.2 Cross-Modal Identity Preservation

The textual consistency established at the prompt level serves as the foundation for maintaining character identity across visual and audio modalities. By ensuring that all downstream generation processes receive consistent character descriptions, we establish a unified basis for character representation that propagates through the entire pipeline.

C Face Swapping Method Comparison

This section provides comprehensive analysis of different face swapping methods used in our Decoupled Character Integration pipeline, supporting the technical choices made in the Cinematographic Synthesis Module.

C.1 Comparative Analysis

We conducted comprehensive experiments comparing different face swapping methods (FSGAN **?**, FaceShifter **?**, and SimSwap **?**) for character integration. Table 1 presents quantitative results across multiple metrics that guided our selection of SimSwap for the final system.

Table 1: Comparison of face swapping methods on identity preservation and quality metrics.

Method	ID Similarity↑	LPIPS↓	FID↓	SSIM↑
FSGAN	0.801	0.221	25.7	0.847
FaceShifter	0.845	0.198	22.3	0.871
SimSwap	**0.891**	**0.176**	**19.8**	**0.894**

C.2 Identity Preservation Metrics

We evaluate identity preservation using multiple complementary metrics to ensure robust character consistency:

Identity Similarity (ID Similarity): Measures cosine similarity in the face recognition embedding space using ArcFace **?**:

$$\text{ID\_Sim} = \frac{\mathbf{f}_{\text{src}} \cdot \mathbf{f}_{\text{swap}}}{\|\mathbf{f}_{\text{src}}\| \|\mathbf{f}_{\text{swap}}\|} \tag{1}$$

where $\mathbf{f}_{\text{src}}$ and $\mathbf{f}_{\text{swap}}$ are the face embeddings of the source and swapped faces respectively.

Learned Perceptual Image Patch Similarity (LPIPS): Evaluates perceptual similarity using the metric proposed by Zhang et al. **?**, measuring perceptual differences between images in deep feature space.

Fréchet Inception Distance (FID): Computes the distance between feature distributions of real and generated images using Inception-v3 features **?**, where lower values indicate higher quality.

Structural Similarity Index Measure (SSIM): Measures structural similarity between images based on luminance, contrast, and structure comparisons **?**, with values closer to 1.0 indicating higher similarity.

D Frame-Level Synchronization

This section provides detailed technical specifications for the audio-visual synchronization mechanisms employed in our Talking Face component, ensuring natural lip movements and temporal alignment.

D.1 Audio-Visual Synchronization

Our talking face generation employs Wav2Lip **?** for audio-visual synchronization, which directly maps audio features to lip movements without explicit phoneme extraction. The synchronization process works as follows:

This approach ensures that character dialogue is synchronized with appropriate lip movements while preserving the visual identity established through the face swapping process.

E Experimental Details and Analysis

This section provides comprehensive details supporting the experimental evaluation presented in the main paper, including per-genre performance analysis and computational efficiency considerations.

E.1 Per-Genre Performance Analysis

Table 2 provides detailed performance breakdown across different movie genres, demonstrating our system's adaptability to diverse narrative styles and emotional contexts.

Algorithm 2 Audio-Visual Synchronization

Require: Audio signal $\mathbf{a}$, video frames $\mathbf{V} = [\mathbf{v}_1, ..., \mathbf{v}_T]$
Ensure: Synchronized video $\mathbf{V}^*$
1: Extract audio features: $\mathbf{F}_a = \text{AudioEncoder}(\mathbf{a})$
2: Extract visual features: $\mathbf{F}_v = \text{VisualEncoder}(\mathbf{V})$
3: **for** $t = 1$ to T **do**
4: Align audio-visual features: $\mathbf{f}_{av}^t = \text{Align}(\mathbf{F}_a^t, \mathbf{F}_v^t)$
5: Generate lip region: $\mathbf{l}_t = \text{LipGenerator}(\mathbf{f}_{av}^t)$
6: Composite with original frame: $\mathbf{v}_t^* = \text{Composite}(\mathbf{v}_t, \mathbf{l}_t)$
7: **end for**
8: **return** $\mathbf{V}^* = [\mathbf{v}_1^*, ..., \mathbf{v}_T^*]$

Table 2: Genre-specific performance analysis. Scores represent averages across 20 test cases per genre.

Genre	OC↑	SC↑	AQ↑	MS↑
Romantic Comedy	0.267	0.953	0.612	0.991
Action Sequence	0.245	0.941	0.587	0.985
Drama	0.271	0.956	0.608	0.989
Family Drama	0.253	0.947	0.595	0.987
Suspense Thriller	0.259	0.948	0.598	0.983
Average	0.259	0.949	0.600	0.987

E.2 Computational Efficiency Analysis

Table 3 presents computational requirements for different system components, providing insights into the scalability and resource allocation of our hierarchical approach.

Table 3: Computational efficiency analysis for 5.375-second video generation.

Component	Time (min)
Narrative Synthesis	2.8
Character Generation	0.5
Scene Creation	6.3
Decoupled Character Integration	1.2
Final Assembly	0.5
Total	11.3

F Ablation Study Details

This section provides comprehensive analysis of the ablation studies presented in the main paper, examining the contribution of each major system component to overall performance.

F.1 Individual Agent Contribution Analysis

We systematically ablate each agent in our narrative synthesis module to understand their individual contributions to the overall system performance:

The results demonstrate that the Character Designer contributes most significantly to Subject Consistency, while the Quality Inspector plays a crucial role in maintaining Overall Consistency across the entire generation pipeline.

F.2 Hyperparameter Sensitivity Analysis

We analyze the sensitivity of key hyperparameters in our system to understand the robustness of our approach:

Table 4: Individual agent ablation results showing the contribution of each component.

Configuration	OC↑	SC↑	AQ↑	MS↑
w/o Character Designer	0.235	0.916	0.571	0.979
w/o Script Writer	0.241	0.934	0.583	0.982
w/o Storyteller	0.238	0.928	0.577	0.981
w/o Composer	0.251	0.943	0.592	0.985
w/o Quality Inspector	0.245	0.938	0.570	0.982
Full System	**0.259**	**0.949**	**0.600**	**0.987**

Agent Communication Rounds: We vary the number of communication rounds T in Algorithm 1 to evaluate the convergence behavior:

Table 5: Sensitivity to number of agent communication rounds.

Rounds	OC↑	SC↑	AQ↑	MS↑
1	0.234	0.925	0.573	0.976
3	0.248	0.941	0.589	0.983
5	**0.259**	**0.949**	**0.600**	**0.987**
7	0.257	0.947	0.598	0.986
10	0.255	0.945	0.596	0.985

The results show that 5 communication rounds provide optimal performance, with diminishing returns for additional rounds due to convergence of the agent coordination process.

G Implementation Details

Our implementation employs a distributed architecture designed for efficient hierarchical movie generation. The system operates on NVIDIA A100 GPUs with high-speed SSD storage for intermediate file handling, ensuring smooth data flow throughout the multi-stage pipeline. The software foundation builds upon PyTorch 2.0.1 with CUDA 11.8 for deep learning operations, integrated with Transformers 4.30.2 for LLM coordination and Diffusers 0.18.2 for video generation models. OpenCV 4.8.0 handles video processing tasks while FFmpeg 5.1.2 manages audio-video synchronization operations. The LLM agents utilize Claude-Sonnet-3.5 API with carefully tuned parameters: temperature settings of 0.7 for creative narrative tasks and 0.3 for technical coordination tasks, maximum token limits of 4096 per agent response, and top-p sampling of 0.9 to balance creativity with coherence. Video generation employs HunyuanVideo-13B with custom fine-tuning for character consistency, configured with 50 inference steps for high quality output, guidance scale of 7.5 for strong text adherence, and 512×512 resolution at 24 FPS for optimal performance. The distributed nature of the architecture allows for parallel processing of different pipeline stages while maintaining strict coordination protocols between agents, enabling efficient scaling for longer content generation while preserving the quality and consistency that characterizes our hierarchical approach.

H Limitations and Future Work

Our framework faces several technical challenges. **Narrative complexity**: capturing nuanced emotional storytelling and sophisticated character development remains challenging compared to human screenwriters. **Long-term consistency**: maintaining character coherence across extended sequences with dramatic emotional transitions in feature-length productions requires further investigation. **Cinematographic sophistication**: replicating advanced directorial techniques such as dynamic camera movements and artistic compositions needs enhancement. **Computational efficiency**: high-quality generation presents scalability challenges for long-form content production, particularly when real-time adjustments are needed. Future work will address these limitations through enhanced narrative modeling, improved long-range dependency mechanisms, and more efficient generation architectures.