



AI Video Generation Pipeline

Background

Video generation with AI has transformed creative production. What once required teams of editors, motion designers, and sound engineers can now be orchestrated through intelligent pipelines that understand context, timing, and visual coherence.

Companies like Runway, Pika, and others have shown us what's possible. But true AI video generation isn't just about creating clips. It's about building cohesive narratives that seamlessly integrate image generation, video synthesis, audio, voiceovers, and timing.

Consider how Midjourney transformed image creation. Now imagine that same revolution for video production. A single prompt generates a complete music video synced to beats, or an entire ad campaign tailored to a brand's visual identity.

This project challenges you to build an end-to-end AI video generation pipeline that creates professional-quality video content with minimal human intervention.

Why This Matters

The future of content creation is generative. Brands need hundreds of ad variations. Musicians want instant music videos. Creators need content at scale.

The team that builds the most robust, cost-effective pipeline wins not just this competition, but potentially defines the future of AI video production. You'll be building technology that could power the next generation of creative tools.

Project Overview

This is a one-week sprint with a **\$5,000 bounty** for the winning team.

Key Deadlines:

- Start: Friday, Nov 14, 2025
- MVP: Sunday (48 Hours)
- Early Submission: Wednesday (5 days)
- Final: Saturday (8 days)

You'll build a complete AI video generation pipeline that takes high-level prompts and outputs publication-ready video content with synchronized audio, coherent visuals, and professional polish.

MVP Requirements (48 Hours)

This is a hard gate. To pass the MVP checkpoint, you must have:

1. Working video generation for at least ONE category (music video OR ad creative)
2. Basic prompt to video flow (text input to video output)
3. Audio visual sync (video matches audio timing/beats)
4. Multi clip composition (at least 3 to 5 clips stitched together)
5. Consistent visual style across clips
6. Deployed pipeline (API or web interface)
7. Sample outputs (at least 2 generated videos demonstrating capability)

The MVP proves your pipeline works end to end. A simple but reliable music video generator beats a feature-rich system that produces incoherent output.

Example MVP Architecture

At minimum, you should have:

1. Prompt Parser: Interprets user input and extracts creative direction
2. Content Planner: Breaks video into scenes/segments with timing
3. Generation Engine: Calls AI models (video, image, audio) for each segment
4. Composition Layer: Stitches clips with transitions and audio sync
5. Output Handler: Renders final video in standard format (MP4, WebM)

Core Pipeline Requirements

Video Categories

You must support at least ONE of these categories with full end to end generation:

Category 1: Music Video Pipeline

Input: Song file (generated or uploaded) + creative direction

Output: Complete music video (1 to 3 minutes)

Requirements:

- Generate or accept AI generated music (Suno, Udio, etc.)

- Analyze song structure (intro, verse, chorus, bridge, outro)
- Detect beats and tempo for scene transitions
- Generate visuals that match song mood and lyrics
- Sync visual transitions to musical beats
- Maintain visual coherence across scenes
- Apply consistent style/aesthetic throughout

Example Prompts:

- "Create an ethereal music video for this ambient electronic track with floating geometric shapes"
- "Generate a high energy punk rock video with urban graffiti aesthetics"
- "Make a dreamy indie pop video with pastel colors and nature scenes"

Category 2: Ad Creative Pipeline

Input: Product description + brand guidelines + ad specifications

Output: Video advertisement (15 to 60 seconds)

Requirements:

- Generate product showcase clips
- Apply brand colors and visual identity
- Create multiple ad variations (A/B testing)
- Support different aspect ratios (16:9, 9:16, 1:1)
- Add text overlays (product name, CTA, price)
- Generate background music or sound effects
- Include voiceover capability (optional but bonus)

Example Prompts:

- "Create a 30 second Instagram ad for luxury watches with elegant gold aesthetics"
- "Generate 3 variations of a TikTok ad for energy drinks with extreme sports footage"
- "Make a product showcase video for minimalist skincare brand with clean white backgrounds"

Category 3: Educational/Explainer Pipeline (Bonus Category)

Input: Topic/script + visual style preferences

Output: Explainer video with narration and visuals

Requirements:

- Generate narration/voiceover from script
- Create visualizations matching narration timing
- Add text captions and graphics
- Maintain educational clarity
- Support diagrams, charts, and animations

Technical Requirements

1. Generation Quality

Visual Coherence:

- Consistent art style across all clips
- Smooth transitions between scenes
- No jarring style shifts or artifacts
- Professional color grading

Audio Visual Sync:

- Beat matched transitions (music videos)
- Voiceover timing (ad creatives)
- Sound effects aligned with visuals
- No audio video drift

Output Quality:

- Minimum 1080p resolution
- 30+ FPS
- Clean audio (no distortion or clipping)
- Proper compression (reasonable file size)

2. Pipeline Performance

Speed Targets:

- 30 second video: Generate in under 5 minutes
- 60 second video: Generate in under 10 minutes
- 3 minute video: Generate in under 20 minutes

Note: We understand AI model inference takes time. We're measuring end to end pipeline efficiency, including smart caching and optimization strategies.

Cost Efficiency:

- Track and report generation cost per video
- Optimize API calls (avoid redundant generations)
- Implement caching for repeated elements
- Target: Under \$2.00 per minute of final video

Reliability:

- 90%+ successful generation rate
- Graceful failure handling
- Automatic retry logic for failed API calls
- Error logging and debugging support

3. User Experience

Input Flexibility:

- Natural language prompts
- Optional parameter controls (style, duration, mood)
- Reference image/video uploads (style transfer)
- Brand guideline documents (for ads)

Output Control:

- Preview generation before final render
- Regenerate specific scenes
- Adjust timing and transitions
- Export in multiple formats

Feedback Loop:

- Show generation progress
- Display which stage is processing
- Preview intermediate results
- Allow user intervention/correction

Advanced Features (Competitive Advantages)

These aren't required but will significantly strengthen your submission:

Style Consistency Engine

- Train custom LoRA models for brand consistency
- Character consistency across scenes
- Automatic style transfer from reference images

Intelligent Scene Planning

- Analyze music structure (AI powered beat detection)
- Generate storyboards before video creation
- Shot variety logic (close ups, wide shots, transitions)

Multi Modal Generation

- Combined image + video generation (static + motion)
- Text to speech with emotion control
- Sound effect generation matching visuals

Iterative Refinement

- Chat interface for video editing
- "Make this scene brighter"
- "Add more motion to the chorus"
- "Change the color palette to warmer tones"

Batch Generation

- Generate multiple variations simultaneously
- A/B testing for ad creatives
- Different aspect ratios from single prompt

Evaluation Criteria

Your pipeline will be judged on these weighted factors:

1. Output Quality (40%)

- Visual coherence: Does it look professional?
- Audio visual sync: Are transitions timed properly?
- Creative execution: Does it match the prompt?
- Technical polish: Resolution, frame rate, compression

2. Pipeline Architecture (25%)

- Code quality: Clean, maintainable, documented
- System design: Scalable and modular
- Error handling: Robust failure recovery
- Performance optimization: Fast and efficient

3. Cost Effectiveness (20%)

- Generation cost: Price per video produced
- API efficiency: Smart caching and optimization
- Resource usage: Memory, compute, storage

4. User Experience (15%)

- Ease of use: Intuitive interface
- Prompt flexibility: Handles varied inputs
- Feedback quality: Clear progress indicators
- Output control: Fine tuning capabilities

Testing Scenarios

We'll evaluate your pipeline with prompts like:

Music Videos:

- "Generate a music video for [attached song] with cyberpunk aesthetics"
- "Create a lo fi hip hop video with cozy study room vibes"
- "Make an epic orchestral video with fantasy landscapes"

Ad Creatives:

- "Create 3 variations of a 15 second Instagram ad for [product description]"
- "Generate a luxury brand video ad with minimal aesthetic"
- "Make a dynamic product showcase for tech gadgets"

Stress Tests:

- Multiple concurrent generation requests
- Very long videos (3+ minutes)
- Complex multi part narratives
- Unusual style combinations

Technical Stack

You'll have access to all the latest image and video generation models on [Replicate](#).

Important: Start development with cheaper models to iterate quickly and control costs. As you approach the showcase, switch to more expensive, higher quality models for your final outputs.

Use whatever stack produces the best results. We care about output quality, not tech stack choices.

Submission Requirements

Submit by Sunday 10:59 PM CT:

1. GitHub Repository

- README with setup instructions and architecture overview
- Documentation explaining pipeline stages
- Cost analysis (breakdown of generation costs)
- Deployed link (API endpoint or web interface)

2. Demo Video (5 to 7 minutes)

Show:

- Live generation from prompt to final video
- Walkthrough of your pipeline architecture
- Comparison of different prompts/styles
- Challenges you solved and trade offs you made

3. AI Generated Video Samples

You must submit at least 3 AI generated videos for your chosen category:

For Music Videos:

- One video synced to an upbeat/energetic song
- One video synced to a slow/emotional song
- One video demonstrating complex visual transitions

For Ad Creatives:

- Three different product ads showing style variation
- At least one ad in vertical format (9:16) for social media
- At least one ad with text overlays and call to action

For Educational/Explainer:

- One technical explanation with diagrams
- One narrative driven explainer
- One demonstration with step by step visuals

4. Technical Deep Dive (1 page)

Answer these questions:

- How do you ensure visual coherence across clips?
- How do you handle audio visual synchronization?
- What's your cost optimization strategy?
- How do you handle generation failures?
- What makes your pipeline better than others?

5. Live Deployment

- Public URL for testing your pipeline
- API documentation if applicable
- Test credentials for judges to access
- Rate limits clearly communicated

Judging Process

Round 1: Initial Review

All submissions reviewed for completeness and basic functionality.

Round 2: Technical Evaluation

Deep dive into code quality, architecture, and innovation.

Round 3: Output Testing

Judges generate videos with standardized prompts and evaluate quality.

Round 4: Final Scoring

Weighted scores across all criteria determine the winner.

Winner Announcement: Monday following submission deadline

Prize Structure

Grand Prize: \$5,000

Best overall video generation pipeline. Combination of quality, cost efficiency, and innovation.

Bonus Recognition:

- Most cost efficient pipeline
- Best music video generator
- Best ad creative generator
- Most innovative architecture

Inspiration

Study these to understand the state of the art:

Companies:

- Runway ML (Gen 3)
- Pika Labs
- Kaiber AI
- Synthesia
- HeyGen
- Kling AI

Concepts:

- Icon's rapid creative generation
- Midjourney's consistent style system
- Modern ad tech platforms (Meta Ads, Google Ads creative studios)

Think about:

- How do professional video editors build music videos?
- What makes an ad creative effective vs generic?
- How do you maintain visual coherence without human oversight?
- What's the minimum viable feature set for real world usage?

Final Note

This is your chance to build technology that could redefine content creation. The best AI video startups are raising millions to solve these exact problems.

A working pipeline that generates ONE category of video beautifully beats a complex system that tries to do everything poorly.

Focus on:

- Coherence over quantity
- Reliability over features
- Cost efficiency over bleeding edge models

Ship something real that actually works.

Let's build the future of video creation.

Questions? Ready to build?

The clock starts now. Make it count.