07pro: The Ultimate Creative Agent Roadmap

Overview

07pro is a multi-modal AI agent designed to transform your imagination, ideas, or stories into fully produced, social-media-ready videos. It accepts multiple input types—including text, audio, images, and even video—and generates a complete package with a branded 3-second intro, a 10–20 second hook, a catchy title, and a custom thumbnail.

Input Modalities

- **Text Input:**
  Enter a prompt, idea, or story. This is the primary driver for script generation.

- **Audio Input:**
  Record your thoughts or narrate ideas. The system can transcribe and use them as a basis for the video.

- **Image Input:**
  Provide inspiration images or reference visuals to set the tone and style.

- **Video Input:**
  Yes, you can use video as an input.

  - **How it Works:**

    - **Analysis & Extraction:** The agent analyzes key frames, detects scene changes, and (if needed) transcribes any dialogue.

    - **Context Building:** It extracts visual themes and narrative cues from the video to enrich the creative process.

Core Capabilities and Workflow

1. **Idea to Script Conversion:**

   - **Llama 4 Integration:**
     Use Llama 4 to generate a detailed, creative script from your prompt. This

model refines your concept, suggesting narrative flow, dialogues, and scene descriptions.

2. **Multimodal Media Generation:**

   ○ **3-Second Channel Intro:**
   Automatically create a short branded clip that introduces your style or channel.

   ○ **10–20 Second Hook:**
   Develop an engaging and viral hook to capture viewers' attention right from the start.

   ○ **Full Video Content:**
   Assemble the main video content using AI-driven video synthesis tools (e.g., Pika Labs, Runway ML, or other text-to-video generators) guided by the script.

   ○ **Thumbnail & Title Creation:**
   Generate a dynamic thumbnail with styled text overlay and craft a click-worthy video title that resonates with social media trends.

   ○ **Voice Cloning & Modification:**
   Incorporate your cloned voice for narration using advanced text-to-speech engines (like ElevenLabs or similar tools), with options to modify tone or style.

3. **Editing & Post-Production:**

   ○ **Video Editing Pipeline:**
   Use automated editing tools (like ffmpeg-based pipelines) to splice together the intro, hook, main content, and overlays into a seamless final video.

   ○ **Multi-Layered Workflow:**
   The agent can process multiple tasks simultaneously—ensuring the creation of intros, hooks, thumbnails, and the final video in a coordinated manner.

Technical Architecture & Integration

● **Frontend & User Interaction:**

   ○ A user-friendly interface (via Hugging Face Spaces, a web app, or even a mobile terminal like Termux) that accepts and previews inputs.

● **Backend Components:**

- ○ **Input Processing Module:**
  Handles video frame extraction, speech-to-text transcription, image recognition, and text parsing.

- ○ **Creative Engine (Llama 4 & Others):**
  Leverages Llama 4 for creative text generation and story expansion, along with other specialized models for voice cloning, image generation, and video synthesis.

- ○ **Media Synthesis & Editing Module:**
  Integrates APIs for video generation, voice synthesis, and automated editing workflows.

- ○ **Output Composer:**
  Assembles all generated pieces (intro, hook, main video, thumbnail, title) into the final product.

- ● **Deployment & Automation:**

  - ○ **Hugging Face Spaces or Custom Hosting:**
    Deploy the agent as an interactive space or web service.

  - ○ **Scheduled/On-Demand Automation:**
    Enable scheduled tasks for continuous content generation and allow real-time, prompt-based creative sessions.

Roadmap Summary Prompt Example

You can use the following text when explaining your agent's concept to developers, collaborators, or even as part of your project documentation:

Agent Name: 07pro
Purpose:
07pro is a multi-modal, AI-driven content creation agent that transforms my ideas into complete, social-media-ready videos. It accepts various input types—including text, audio, image, and video—to create a final product that includes:

- ● A 3-second branded channel intro,

- ● A 10–20 second attention-grabbing hook,

- ● A fully scripted and produced video,

- ● A custom thumbnail and viral video title,

- And narration using my cloned and modified voice.

Core Features:

- **Input Versatility:** Supports text, audio, image, and video input (with video analysis for frame extraction and transcription).

- **Creative Script Generation:** Utilizes Llama 4 to craft engaging, detailed scripts based on my ideas.

- **Multi-Modal Media Production:** Automatically generates a branded intro, hook, and main video content using cutting-edge text-to-video and editing technologies.

- **Voice Cloning & Synthesis:** Employs advanced text-to-speech models for authentic narration in my voice.

- **Automated Post-Production:** Combines all elements using video editing pipelines to produce a polished final product.

Workflow:

1. **Input Analysis:** Accept and analyze text, audio, image, or video inputs.

2. **Script Generation:** Leverage Llama 4 for creative scriptwriting.

3. **Media Generation:** Create a 3s intro, 10–20s hook, and the main video.

4. **Voice Integration:** Generate narration using cloned voice technology.

5. **Final Assembly:** Automatically edit and assemble all components (intro, hook, video, thumbnail, title) into one cohesive video ready for social media.

This detailed roadmap and prompt provide a comprehensive blueprint for your agent 07pro, covering every aspect from multi-modal inputs (including video) to output production and technical integration with Llama 4. This should serve well to explain your vision to anyone asking about the kind of agent you're creating.