**1. Complete Model Con**fi**guration with Fallbacks**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **1.1 Script Generation ModelsTier** | **Provider** | **Primary Model** | **Cost/1Ktokens** | **Fallback Model** | **Cost/1Ktokens** | **Fallback 2** |
| **Free** | OpenRouter | arliai/qwq-32b-arliai-rpr-v1:free | $0.00 | meta-llama/llama-3.2-3b-instruct | $0.00006 | mistralai/mistral-7b-instruct |
| **Basic** | OpenRouter | deepseek/deepseek-chat-v3 | $0.00014 | mistralai/mistral-7b-instruct | $0.00006 | meta-llama/llama-3.2-3b |
| **Standard** | OpenRouter | anthropic/claude-3-haiku | $0.00025 | openai/gpt-3.5-turbo | $0.0005 | deepseek/deepseek-chat-v3 |
| **Pro** | OpenRouter | openai/gpt-4o-mini | $0.00015 | anthropic/claude-3-haiku | $0.00025 | openai/gpt-3.5-turbo |
| **Premium** | OpenRouter | openai/gpt-4o | $0.0025 | anthropic/claude-3-opus | $0.015 | openai/gpt-4o-mini |
| **Enterprise** | Direct API | openai/gpt-4o | $0.0025 | anthropic/claude-3-opus | $0.015 | Custom ensemble |

**1.2 Image Generation Models (Scene Creation)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tier** | **PrimaryProvider** | **Primary Model** | **FallbackProvider** | **Fallback Model** | **Fallback 2** | **Est.Cost/Image** |
| **Free** | ModelsLab | Google/Imagen 4Fast | ModelsLab | flux-dev | fal-ai/qwen-image-v4 | $0.002-0.003 |
| **Basic** | ModelsLab | Modelslab/FLUXText-to-Image | ModelsLab /Fal.ai | Google/Imagen 4Ultra | fal-ai/nano-banana | $0.004-0.006 |
| **Standard** | ModelsLab | Google/Imagen 4Ultra | BytePlus | seedream-4-2k | ModelsLabFLUX-pro | $0.008-0.012 |
| **Pro** | BytePlus | seedream-4-2k | GoogleVertex | imagen-4.0-ultra-preview | ModelsLabImagen 4 Ultra | $0.015-0.020 |
| **Premium** | GoogleVertex | imagen-4.0-ultra-preview | Tencent | hunyuan-image-3.0 | fal-ai/nano-banana (Gemini2.5 Flash) | $0.030-0.040 |
| **Enterprise** | MultipleAPIs | hunyuan-image-3.0 /Gemini 2.5 Flash | GoogleVertex | imagen-4.0-ultra-preview | Customensemble | $0.040-0.080 |

**1.3 Audio Generation Models (Voice/Narration)**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Tier** | **PrimaryProvider** | **Model** | **Cost/1Kchars** | **FallbackProvider** | **Fallback Model** | **Fallback 2** |
| **Free** | ModelsLab | ElevenLabs-clone(basic) | ~$0.001 | OpenAI | tts-1 | ModelsLab basic-tts |
| **Basic** | ModelsLab | ElevenLabs v2 | ~$0.003 | OpenAI | tts-1-hd | ModelsLabElevenLabs-clone |
| **Standard** | ElevenLabsDirect | eleven\_multilingual\_v2 | ~$0.015 | OpenAI | tts-1-hd | ElevenLabs turbo |
| **Pro** | ElevenLabsDirect | eleven\_turbo\_v2\_5 | ~$0.015 | ElevenLabs | eleven\_flash\_v2\_5 | OpenAI tts-1-hd |

**1.4 Video Generation Models**

**A. NARRATION VIDEOS (Audioless - Image-to-Video + Separate Audio Merge)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Tier** | **PrimaryModel** | **APIProvider** | **Resolution** | **Cost/sec** | **Fallback Model** | **FallbackProvider** | **Fallback 2** |
| **Free** | wan2.2-i2v-flash | ModelsLab | 480p | $0.015 | byteplus/seedance-i2v | ModelsLab | wanx2.1-i2v-turbo |
| **Basic** | modelslab-google/veo2-i2v | ModelsLab | 720p | $0.036 | byteplus/seedance-i2v | ModelsLab | wan2.2-i2v-flash |
| **Standard** | veo-3-i2v | GoogleVertex | 720p | $0.50 | veo-3-fast-i2v | GoogleVertex | wan2.2-i2v-plus |
| **Pro** | veo-3-fast-i2v | GoogleVertex | 1080p | $0.40 | seedance-1.0-pro-i2v | BytePlus | veo-3-i2v |
| **Premium** | sora-2-i2v | OpenAIAPI | 1080p | $0.50 | sora-2-pro-i2v | OpenAI API | veo-3-fast-i2v |
| **Enterprise** | sora-2-pro-i2v | OpenAIAPI | 1080p | $1.00 | ray3-i2v-hdr | Luma AI | seedance-1.0-pro-i2v |

**B. CHARACTER DIALOGUE (Audioed - Text-to-Video with Native Audio)**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Tier** | **PrimaryModel** | **APIProvider** | **Resolution** | **Cost/sec** | **Fallback Model** | **FallbackProvider** | **Fallback 2** |
| **Free** | veo-3-fast-preview-t2v | ModelsLab | 720p | $0.40 | veo-3-fast-t2v | GoogleVertex | wanx2.5-t2v-audio |
| **Basic** | veo-3-fast-t2v | ModelsLab | 720p | $0.40 | veo-3-t2v | GoogleVertex | wanx2.5-t2v-audio |
| **Standard** | veo-3-t2v | GoogleVertex | 720p | $0.75 | wanx2.5-t2v-audio | AlibabaCloud | veo-3-fast-t2v |
| **Pro** | seedance-1.0-pro-t2v | BytePlus | 1080p | ~$0.20 | veo-3-fast-audio | GoogleVertex | wanx2.5-t2v-audio |
| **Premium** | veo-3-fast-audio | GoogleVertex | 1080p | $0.40 | veo-3-audio | GoogleVertex | seedance-1.0-pro-t2v |
| **Enterprise** | veo-3-audio | GoogleVertex | 1080p | $0.75 | sora-2-t2v(when available) | OpenAI | seedance-1.0-pro-t2v |

Generation Fallback Chain (3 attempts before failure):

:NARRATION VIDEOS (Image-to-Video):

: Attempt 1Attempt 1:: Primary Model (tier --specific)

└─ \_Wait:: 30 --60 result

Attempt 2Attempt 2:: Fallback Model (if primary fails)

└─ \_Wait:: 45 --90 result

Attempt 3Attempt 3:: Fallback 2 (if Fallback fails)

└─ \_Wait:: 30 seconds for result

Attempt 4Attempt 4:: Error

DIALOGUE VIDEOS (Text-to-Video with Audio)::

Attempt 1Attempt 1:: Primary Model (tier --specific)

└─ \_Wait:: 45 --120 result120 seconds for result

Attempt 2Attempt 2:: Fallback Model (if primary fails)

└─ \_Wait:: 60 seconds for result

Attempt 3Attempt 3:: Fallback 2 ((if Fallback fails)

└─ \_Wait:: 60 seconds for result

Attempt 4Attempt 4 Error

SCRIPT GENERATIONSCRIPT GENERATION::

Attempt 1Attempt 1:: Primary Model

Attempt 2Attempt 2:: Fallback Model

Attempt 3Attempt 3:: Fallback 2

Attempt 4Attempt 4:: Return cached/template script + notify user

IMAGE GENERATIONIMAGE GENERATION::

Attempt 1Attempt 1:: Primary Model

Attempt 2Attempt 2:: Fallback Model

Attempt 3Attempt 3:: Fallback 2

Attempt 4Attempt 4:: Error

AUDIO GENERATIONAUDIO GENERATION::

Attempt 1Attempt 1:: Primary Model

Attempt 2Attempt 2:: Fallback Model

Attempt 3Attempt 3:: Fallback 2

Attempt 4Attempt 4:: Error