Character Consistency

Visual Identity and State Maintenance in Agentic AI

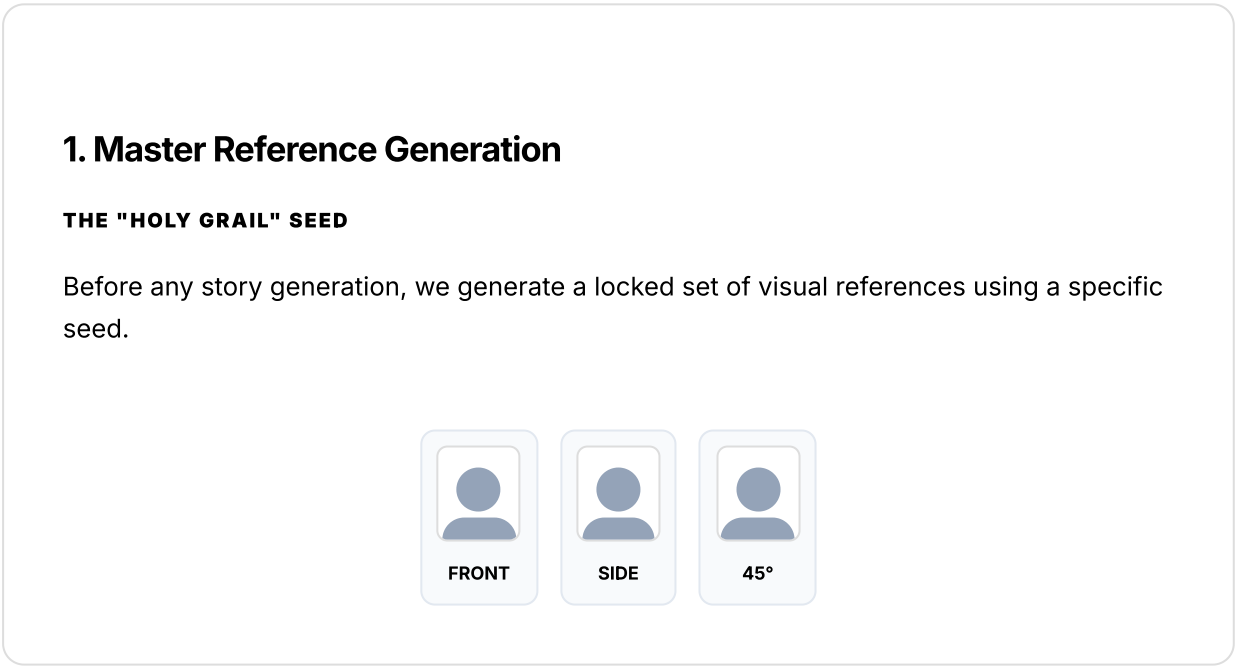
The primary failure mode of long-form AI video generation is **“Character Drift.”** When a model generates a character across multiple scenes, subtle changes in facial structure, hair color, or lighting can destroy narrative immersion. Our Character Consistency Maintenance System (CCMS) is the architectural subsystem designed to prevent this by separating visual identity from narrative action.

Character Consistency Maintenance System (CCMS)

The CCMS operates in three distinct phases, moving from raw pixel data to mathematical embeddings that guide the final output.

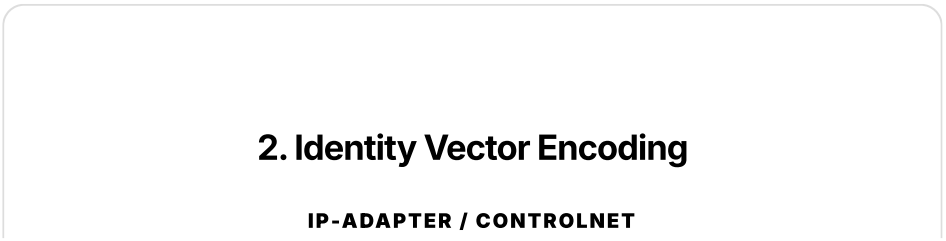
Phase 1: Master Reference Generation

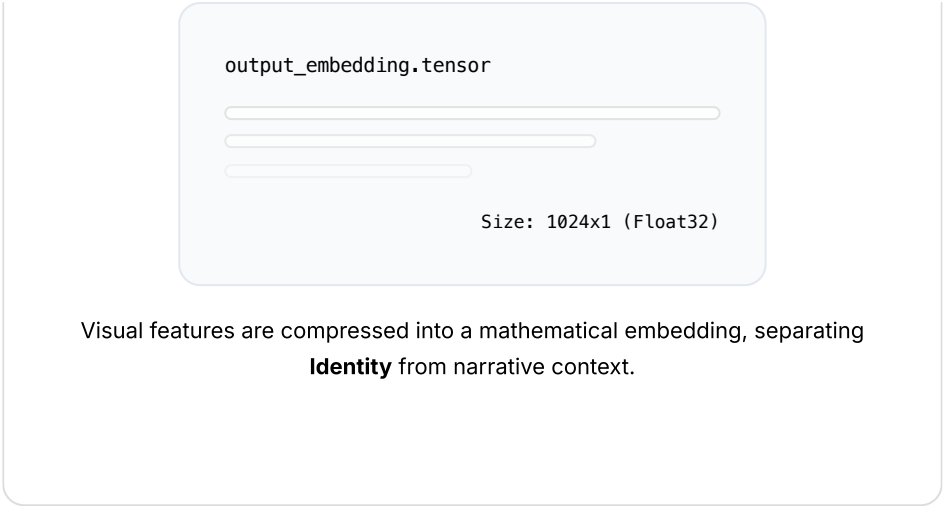
Before any story generation, we generate a locked set of visual references using a specific seed. The system creates a “Holy Grail” reference sheet for each character, capturing multiple angles and expressions.



Phase 2: Identity Vector Encoding

These images are not used directly as prompts. Instead, they are passed through an image encoder (like IP-Adapter or ControlNet reference) to create a visual embedding—a mathematical “Identity Vector.” This separates the character’s *identity* from the *environment* or *action*.

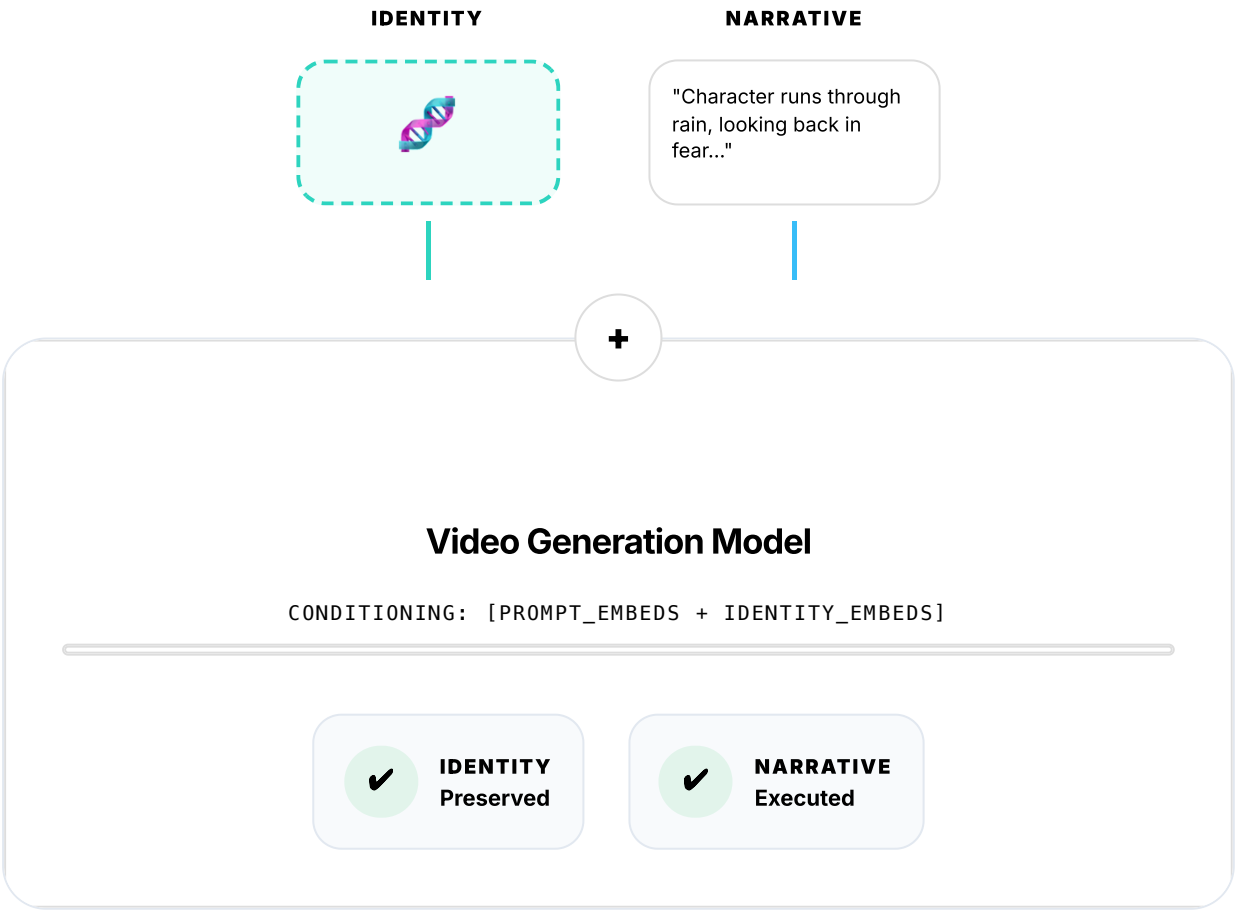




Visual features are compressed into a mathematical embedding, separating **Identity** from narrative context.

Phase 3: Injection and Narrative Synthesis

For every single video generation task, this embedding is passed alongside the text prompt. The video model is instructed to generate the action described in the text, but to force the visual features to match the identity embedding.



The Outfit Manager

In a novel, characters change clothes. The CCMS tracks **"Current Outfit State"** as a discrete variable.

- The **Continuum Flow** engine tracks narrative time. When the text says "He donned his armor," the State Manager updates the `Current_Outfit` variable for that character ID in the backbone.
- Subsequent video prompts automatically inject the `armor_embedding` instead of the `casual_clothes_embedding` without the text chunk explicitly mentioning armor every time.

> *"This decoupling allows the narrative agent to focus on what is happening, while the CCMS ensures who it is happening to remains visually immutable."*

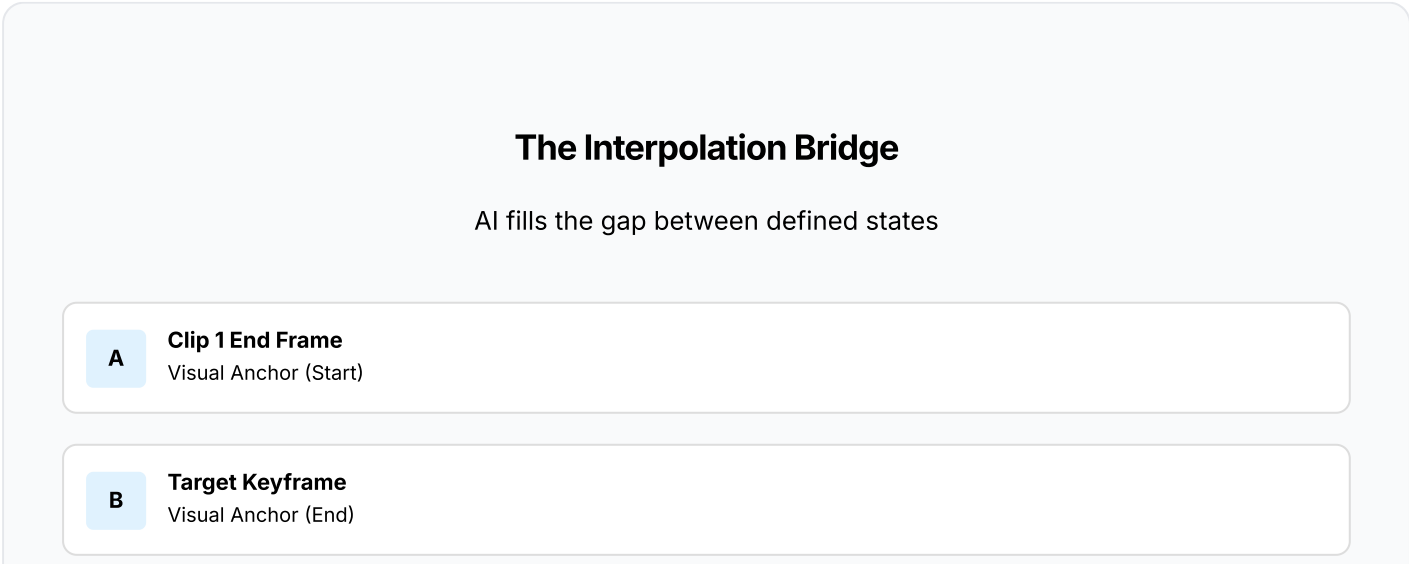
Phase 4: Temporal Continuity (Frame Interpolation)

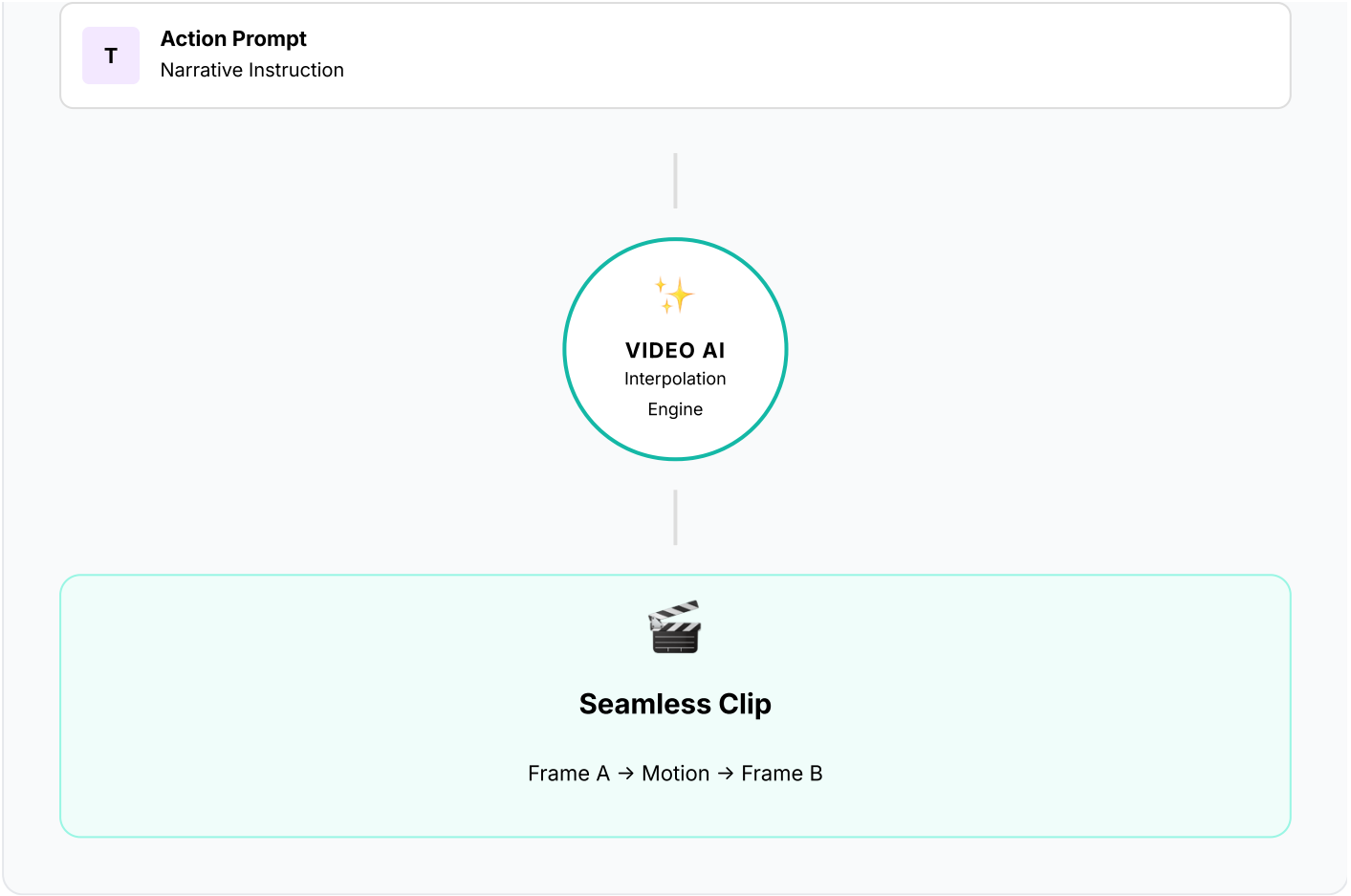
Most modern video generation applications no longer rely on a single image prompt. Instead, they use a **"First Frame & Last Frame"** technique to ensure seamless consistency.

The "Bridge" Strategy

The video generator acts as an interpolation engine, filling in the movement between two known states.

1. **Frame A (Start)**: The actual last frame of the *previous* video clip.
2. **Frame B (End)**: A newly generated "Keyframe" representing the target state.
3. **Generation**: The AI generates the transformation derived from the text prompt.





This ensures that the character doesn’t just “look similar” but is pixels-perfectly continuous from the previous shot.