

Clinical review & integration

Input

Processing

Output

Text Prompt

*“Nurturing adult
providing safety to a
vulnerable child in a
calm outdoor setting”*

Prompt
Refinement
NLP Rephrasing

Prompt Encoder
(e.g., T5, PaLM 2)
Instruction-Tuned Language Model

Text Embedding

Personalization / Emotion
Control Module

Emotion Context Encoding | Clinical & Personal Context Encoding

✓ Emotion Embeddings
e.g., EmoCLIP, EmotiBert
✓ Affective Vectors:
mood, intensity vectors
✓ Control Adapters:
e.g., ControlNet, LoRA

✓ Demographics (age, history)
✓ Therapeutic goals
✓ Environment Preferences

Emotionally-Conditioned
Embeddings

Personalized
Embeddings

Image Input



Therapist-approved visual anchor
(symbolic avatar, abstract child form, or
photo if clinically appropriate)

Image Encoder
(e.g., CLIP, SigLIP)
Visual Identity Encoder

Image Embeddings

Cross-Modal Embedding
Fusion Layer
e.g., BLIP-2, Flamingo, Q-former

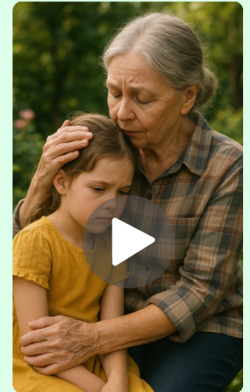
Joint Latent Representation
(text and image)

Generative Video Backbone

✓ Transformer-Based Models
Narrative Conditioning Module
✓ Latent Diffusion
Video Synthesis Core
✓ Cascaded Latent Diffusion
Temporal Refinement for Coherence

5-10 s clip

Emotionally Aligned
Short Video



*All personal data handled with emotional and privacy safeguards

Primary Actor: Therapist

- ✓Generate Prompts
- ✓Adjust emotional tone or script directions
- ✓Guide narrative pacing & safety boundaries