

# Real-Time Video Generation for Holographic Interactions with Bella

## Technical Report

### Abstract

Project Bella focuses on creating an interactive holographic fish companion for children in sterile hospital rooms. This report outlines the technical approach, findings, and solutions developed by the AI Clinic team in partnership with IMAGE et CONCEPT. Our research explores state-of-the-art video generation methods, evaluates their feasibility, and the methods needed to implement a hybrid system to deliver engaging, emotionally adaptive interactions.

## 1 Project Overview

### 1.1 Background

IMAGE et CONCEPT, a French audiovisual solutions leader, developed 'VID,' a volumetric holographic device to enhance the experience of children in sterile hospital environments. Renowned for collaborations with Carrefour, Dior, and France24, the company embarked on this mission to merge cutting-edge technology with emotional care, targeting improved well-being and communication in healthcare settings.

### 1.2 Objectives

- Develop a real-time interactive holographic companion (Bella).
- Enable natural, emotionally adaptive interactions with the children.
- Synchronize animations with audio and emotional cues for immersive engagement.

### 1.3 Impact Goals

- Humanize the hospital experience for children by providing companionship.
- Facilitate communication between children in isolating environments.
- Offer emotional and psychological support tailored to each child's state.

## 2 Research and Development

### 2.1 Research Tracks

#### State-of-the-Art Models:

- **Open Source:**
  - Cog Video
  - Pyramid Flow
  - AnimateLCM
- **Paid Models:**
  - Runway
  - Deepmotion
  - LUMALABS
  - HailuoAI
  - Synthesia
  - KlingAI

#### Research Efforts:

1. API Solutions: Investigated commercial video generation APIs for real-time performance. Assessed cost, compatibility, and scalability for integration with holographic hardware.
2. Open Source Tools: Explored advanced free models like Pyramid Flow and Cog Video. Evaluated their ability to meet real-time video generation requirements.

### 2.2 Findings

#### API Solutions:

- Strengths: Commercial APIs provided polished results and robust infrastructure for video generation.
- Weaknesses(refer to table1 in the end of the report ):
  - High costs rendered them impractical for our resource constraints.
  - Limited customization made it challenging to tailor the outputs for Bella's unique requirements.

#### Open Source Models:

##### 1. Pyramid Flow:

- Strengths: Produced visually appealing outputs, demonstrated potential for generating smooth transitions.

- Weaknesses: Extremely resource-intensive, slower performance unsuitable for real-time interaction.

## 2. Cog Video:

- Strengths: Delivered more efficient performance, flexible integration with Bella's animation pipeline.
- Weaknesses: Challenges in achieving real-time processing speeds, limited animation variety required supplemental pre-rendered clips.

### Work Flow of the Intended Solution:

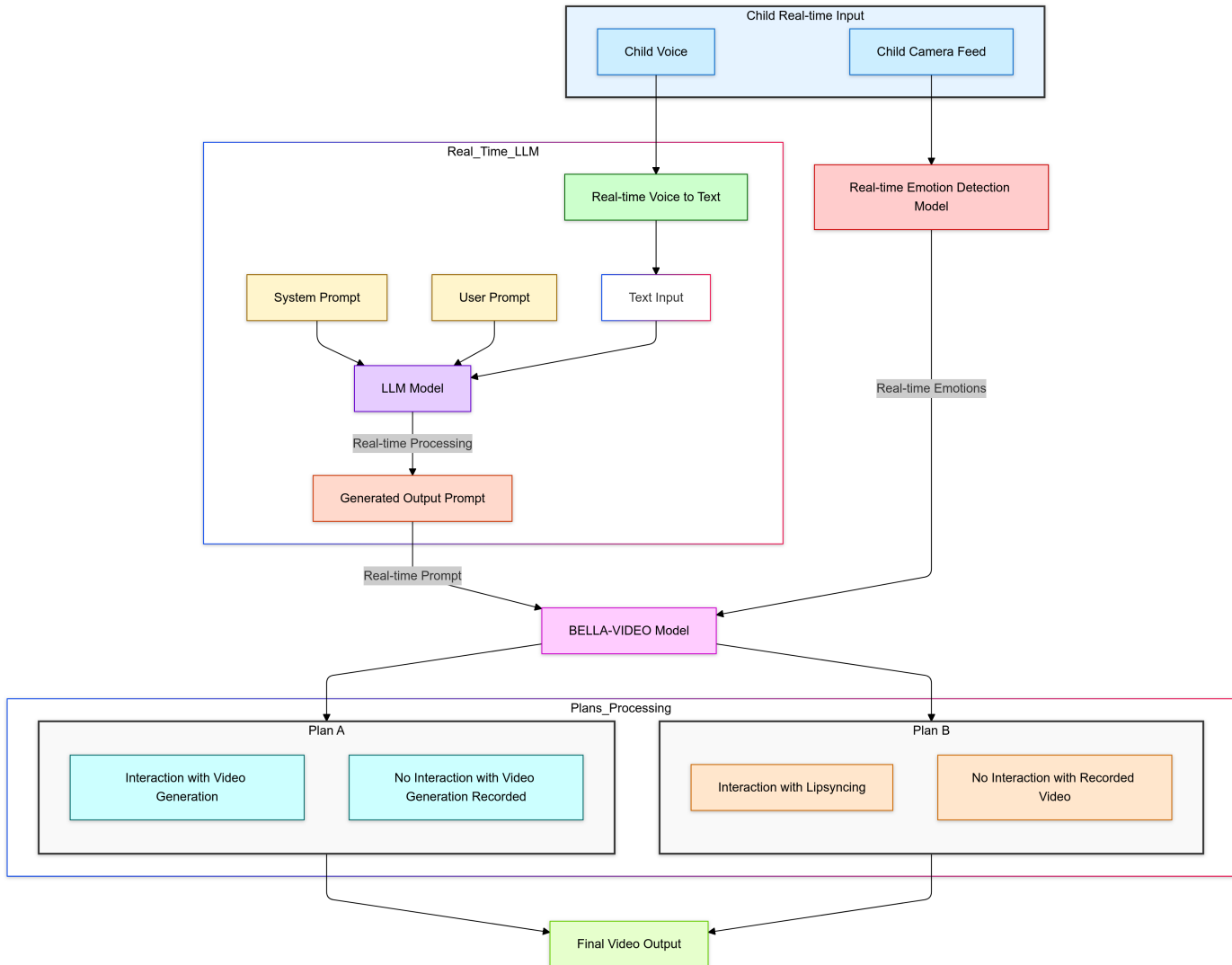


Figure 1: Schema for the Model

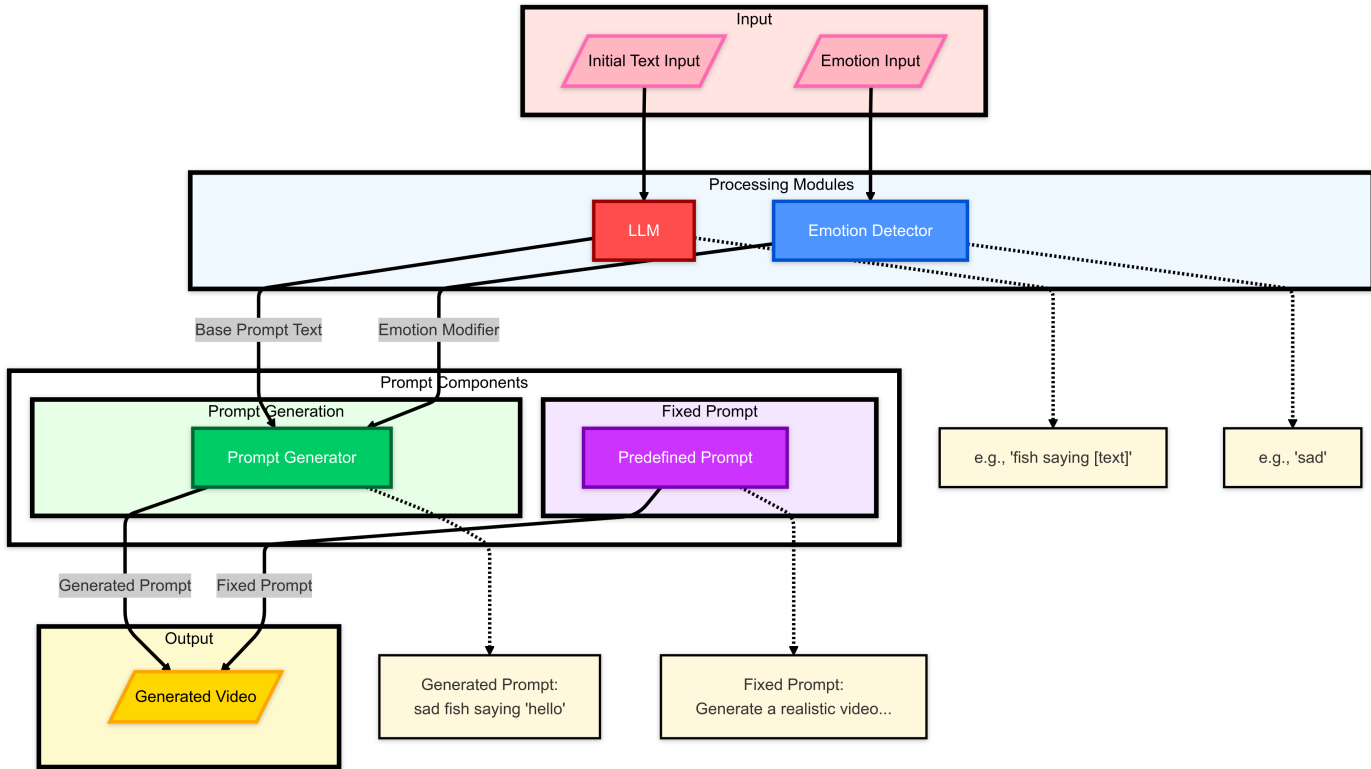


Figure 2: Schema of the Intended Solution

## 3 Challenges and Solutions

### 3.1 Current Challenges

1. Computational and memory demands for real-time generation.
2. Ensuring smooth transitions between clips.
3. Adapting human-centric face detection models for Bella.

### 3.2 Addressed Solutions

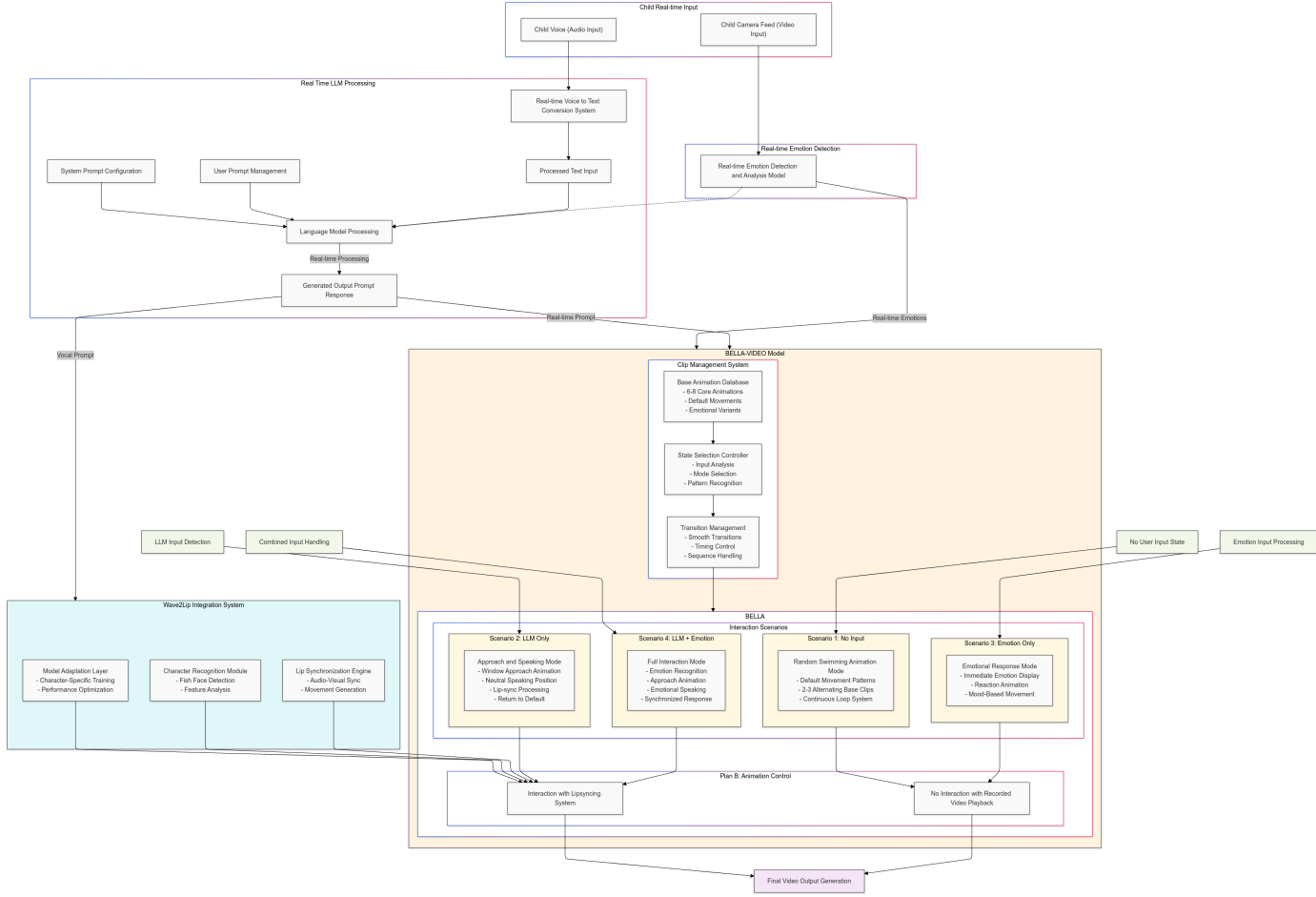


Figure 3: Schema of the Suggested Solution

- Expanded animation sets and standardized transitions.
- Optimized hybrid processing to prioritize system responsiveness.

## 4 Future Development

### 4.1 Short-Term Goals

- Refine lip synchronization.
- Increase animation library diversity.

### 4.2 Long-Term Vision

- Advance real-time video generation with enhanced infrastructure.
- Integrate Bella into hospital communication workflows.
- Expand deployment to educational and therapeutic settings.

## 5 Conclusion

Project Bella demonstrated the potential of holographic companions to humanize sterile hospital environments. While real-time video generation remains a challenge due to resource constraints, the implemented clip-based system, supported by Cog Video and Wave2Lip, delivers engaging and adaptive interactions. Future improvements aim to overcome current limitations and extend the system’s applications.

API Name	Response Time	Cost	Quality	Usage Limits	Pros and Cons
SYNTHESIA		Creator: \$64/mo (billed yearly), Enterprise: Custom pricing		Enterprise: 360 mins/year + paid add-on	<b>Pros:</b> High-quality video output, Reliable API responses, Multi-language support <b>Cons:</b> High cost for larger usage, Limited flexibility on lower-tier plans
LUMA LABS		\$0.4 for 5s at 24fps (1280x720p)	720p resolution, 24fps, high-quality output	5-second video duration at 720p, billed per generated pixel	<b>Pros:</b> Cost-effective for high-quality short videos, Flexible pixel-based pricing <b>Cons:</b> Higher cost for longer or higher-res videos, Limited to short durations
VEO				Not available in France	<b>Pros:</b> Advanced technology with potential for high-quality output <b>Cons:</b> Limited availability, not yet available in France
RUNWAY		5s: \$0.25 per video, 10s: \$0.50 per video	720p resolution with 5s or 10s video outputs	5 or 10 second durations, Gen-3 Alpha Turbo supports up to 34 seconds	<b>Pros:</b> Fast generation times with Turbo model, Flexible text/image to video options, Configurable output settings (aspect ratio, keyframes) <b>Cons:</b> Limited resolution options, Costs increase with longer videos, Requires input image for Turbo model
DEEPMOTION		\$83/month (Studio plan, paid annually)	Up to 4320p (8K) at 240 FPS, high-quality output	Unlimited clips, first 7200 high-priority credits monthly	<b>Pros:</b> Highest quality output, Unlimited animation, Variety of output formats <b>Cons:</b> Higher cost due to annual commitment, Potential overkill for basic projects
KLING AI		\$6720 for 20000 units (20% off, \$0.112 per unit)	Up to 4320p (8K), customizable video modes	20000 units for 3 months, supports 5 concurrent sessions	<b>Pros:</b> High-quality output, Supports multiple sessions, Cost-effective per unit for high volume <b>Cons:</b> Expensive upfront cost, Unused units expire after 3 months

Table 1: Comparison of Video Generation APIs