Real-Time Video Generation for Holographic Interactions with Bella

TABLE OF CONTENTS

01

PROJECT OVERVIEW

PROJECT OVERVIEW

PROJECT : BELLA

- Holographic companion device for hospitalized children
- Real-time interactive technology
- Emotion-adaptive responses

KEY GOALS

- Enhance hospital experience through companionship
- Break isolation barriers
- Provide personalized emotional support

TECHNICAL FOCUS

- Real-time interactivity
- Emotion recognition
- Synchronized audio-visual experiences

02

RESEARCH & DEVELOPMENT

State of the Art Models:

Open Source

- Cog Video
- Pyramid Flow
- AnimateLCM

Paid Models

- Runway
- Deepmotion
- LUMALABS
- HailuoAI
- Synthesia
- KlingAI

Research Efforts:

API Solutions:

- Investigated commercial video generation APIs for real-time performance.
- Assessed cost,compatibility, and scalability for integration with holographic hardware.

Open Source Tools:

- Explored advanced free models like Pyramid Flow and Cog Video.
- Evaluated their ability to meet real-time video generation requirements.

WHAT APIs DID WE FIND?

API Name	COST	PROS	CONS
SYNTHESIA	$64/mo (billed yearly)	High-quality video output	High cost for larger usage, Limited flexibility on lower-tier plans
LUMA LABS	$0.4 for 5s at 24fps (1280x720p)	Flexible pixel- based pricing	Higher cost for longer or higher-res videos, Limited to short durations
VEO		Advanced technology with potential for high-quality output	Limited availability, not yet available in France
RUNWAY	5s: $0.25 per video, 10s: $0.50 per video	Configurable output settings (aspect ratio, keyframes)	Costs increase with longer videos, Requires input image for Turbo model
DEEPMOTION	$83/month (Studio plan, paid annually)	Unlimited animation, Variety of output formats	Higher cost due to annual commitment, Potential overkill for basic projects
KLING AI	$6720 for 20000 units (20% off, $0.112 per unit)	Cost effective per unit for high volume	Expensive upfront cost, Unused units expire after 3 months

03

FINDINGS

FINDINGS : Commercial API Solutions

ADVANTAGES

- Polished video generation
- Ready-to-use infrastructure
- Professional output quality

LIMITATIONS

- Cost-prohibitive
- Restricted customization options
- Poor fit for Bella's specific needs

Key Point: While commercial APIs offered a refined solution, their high costs and inflexibility made them unsuitable for our specialized holographic companion project.

FINDINGS : Open Source Model Comparison

PYRAMID FLOW

Advantages:

- High-quality visual output
- Smooth animation transitions

Limitations:

- Heavy resource consumption
- Performance too slow for real-time use

COG VIDEO

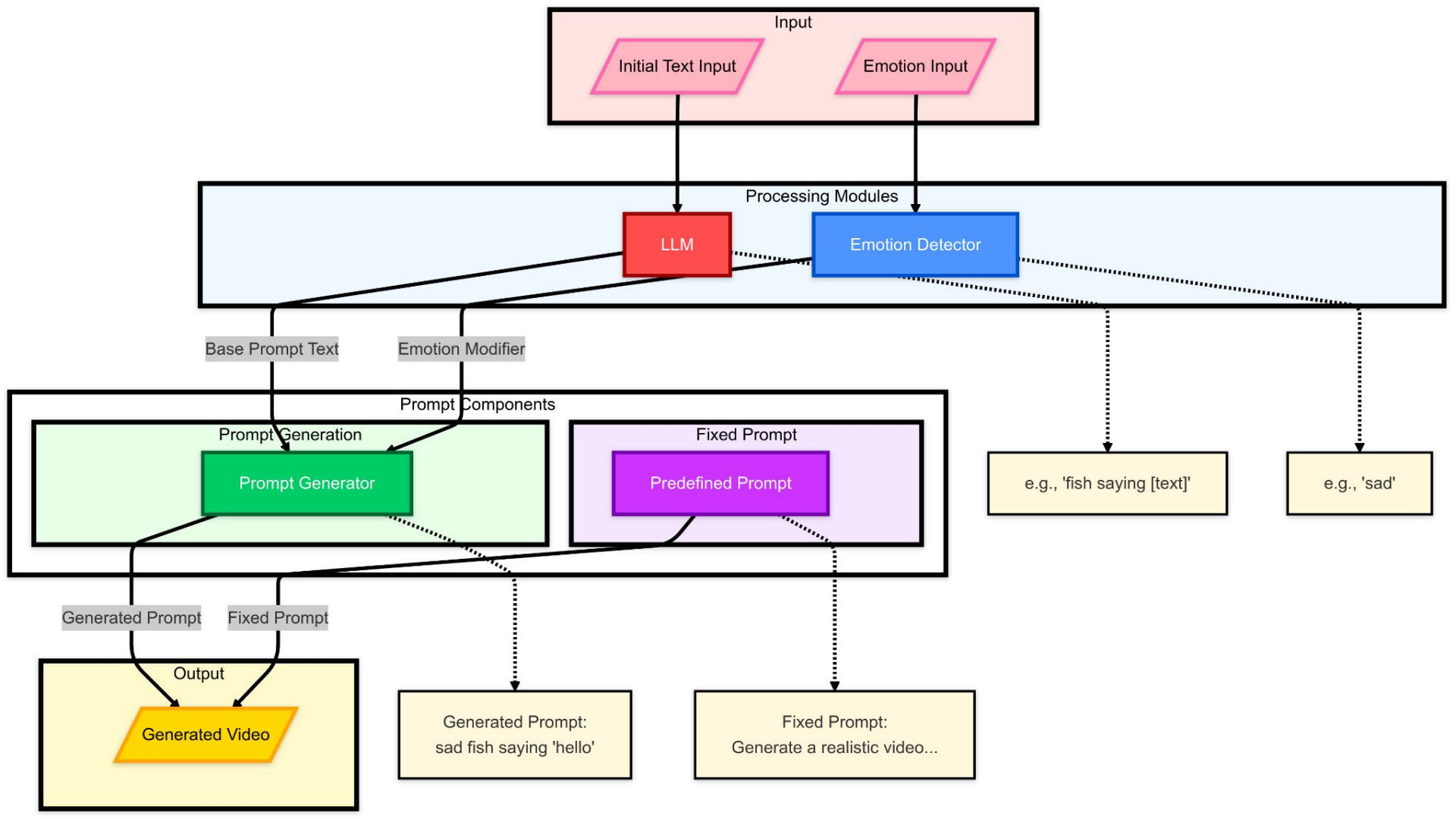
Advantages:

- More efficient processing
- Better pipeline integration

Limitations:

- Still not reaching real-time speeds
- Limited animation range
- Reliance on pre-rendered content

The key takeaway is that while both solutions offered distinct benefits, neither fully achieved the real-time performance needed for seamless interactive holographic


Input
Initial Text Input
Emotion Input
Processing Modules
LLM
Emotion Detector
Base Prompt Text
Emotion Modifier
Prompt Components
Prompt Generation
Prompt Generator
Fixed Prompt
Predefined Prompt
e.g., 'fish saying [text]'
e.g., 'sad'
Generated Prompt
Fixed Prompt
Output
Generated Video
Generated Prompt:
sad fish saying 'hello'
Fixed Prompt:
Generate a realistic video...

04

CHALLENGES & SOLUTIONS

Challenge 1

Computational and memory demands for real-time generation.

Challenge 2

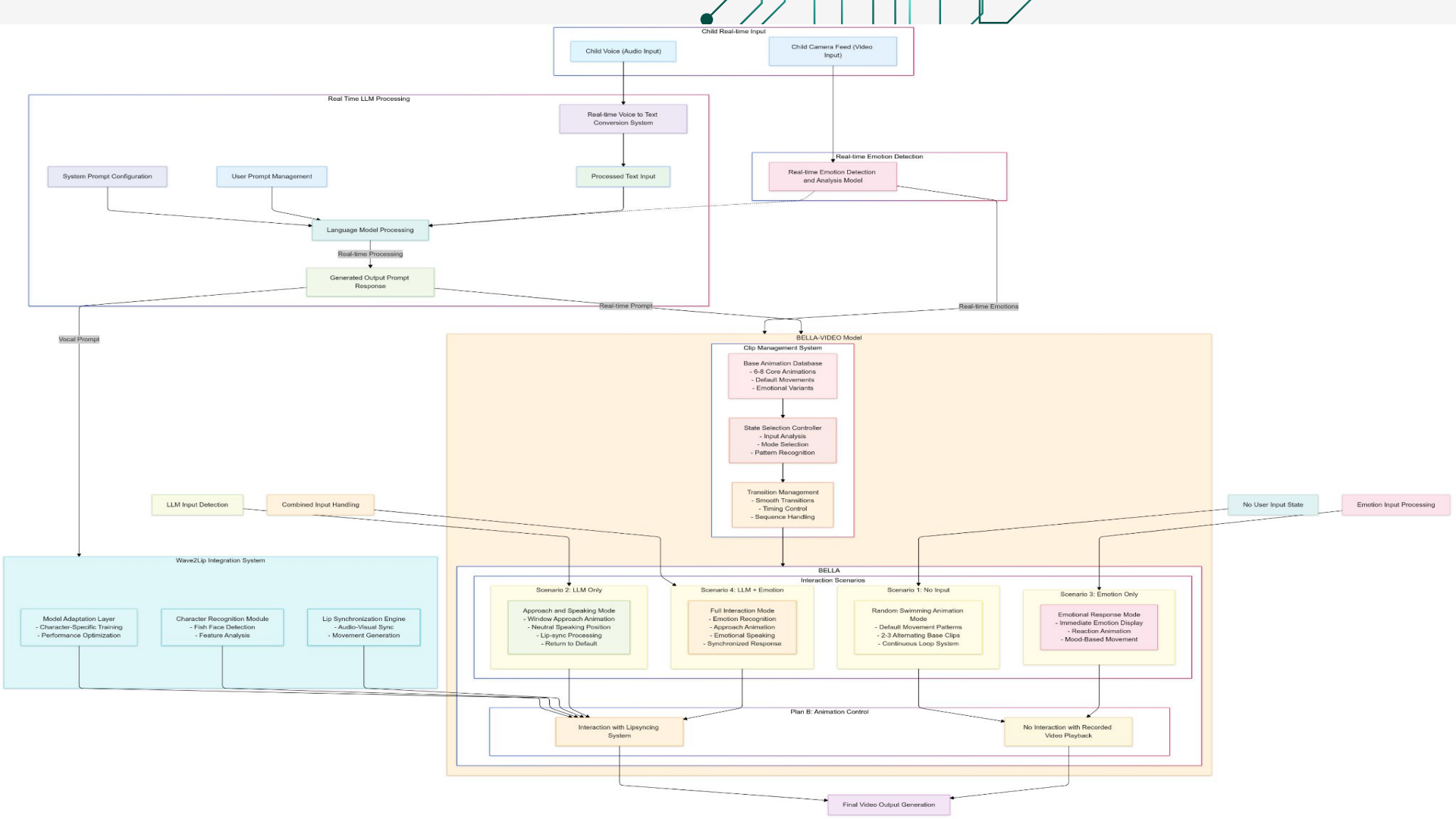
Ensuring smooth transitions between clips.

Challenge 3

Adapting human-centric face detection models for Bella

IMPROVEMENTS

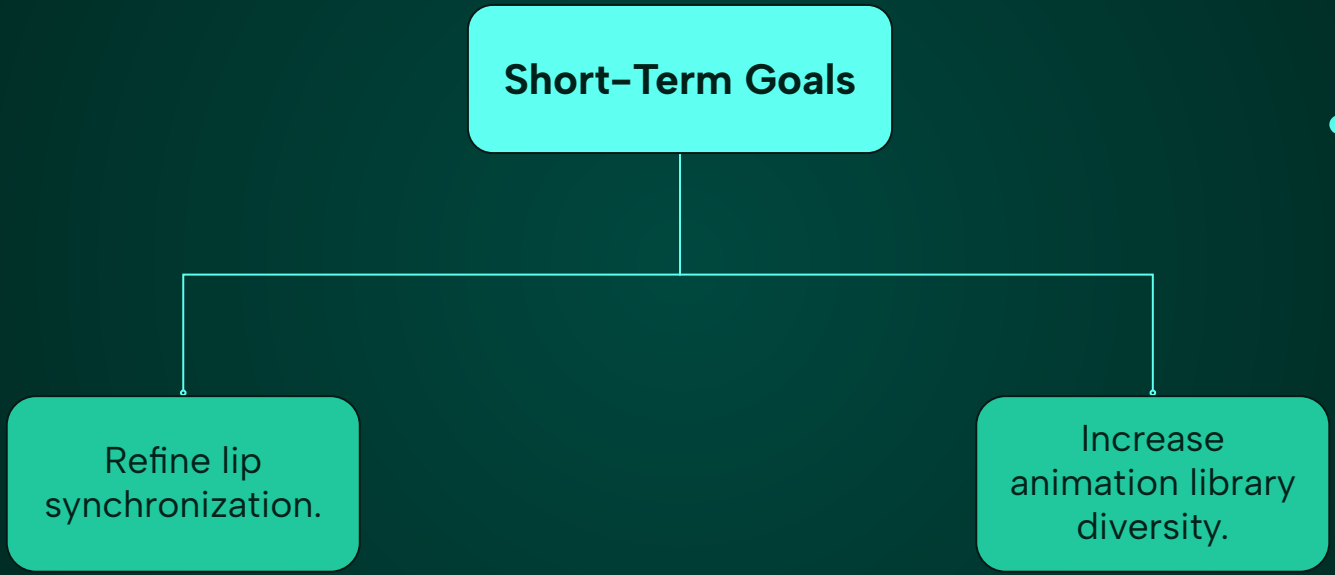
- Expanded animation sets and standardized transitions
- Optimized hybrid processing to prioritize system responsiveness.


Child Real-time Input
Child Voice (Audio Input)
Child Camera Feed (Video Input)
Real Time LLM Processing
Real-time Voice to Text Conversion System
System Prompt Configuration
User Prompt Management
Processed Text Input
Real-time Emotion Detection
Real-time Emotion Detection and Analysis Model
Language Model Processing
Real-time Processing
Generated Output Prompt Response
Real-time Prompt
Real-time Emotions
Vocal Prompt
BELLA-VIDEO Model
Clip Management System
Base Animation Database
- 6-8 Core Animations
- Default Movements
- Emotional Variants
State Selection Controller
- Input Analysis
- Mode Selection
- Pattern Recognition
Transition Management
- Smooth Transitions
- Timing Control
- Sequence Handling
LLM Input Detection
Combined Input Handling
No User Input State
Emotion Input Processing
Wave2Lip Integration System
Model Adaptation Layer
- Character-Specific Training
- Performance Optimization
Character Recognition Module
- Fish Face Detection
- Feature Analysis
Lip Synchronization Engine
- Audio-Visual Sync
- Movement Generation
BELLA
Interaction Scenarios
Scenario 2: LLM Only
Approach and Speaking Mode
- Window Approach Animation
- Neutral Speaking Position
- Lip-sync Processing
- Return to Default
Scenario 4: LLM + Emotion
Full Interaction Mode
- Emotion Recognition
- Approach Animation
- Emotional Speaking
- Synchronized Response
Scenario 1: No Input
Random Swimming Animation Mode
- Default Movement Patterns
- 2-3 Alternating Base Clips
- Continuous Loop System
Scenario 3: Emotion Only
Emotional Response Mode
- Immediate Emotion Display
- Reaction Animation
- Mood-Based Movement
Plan B: Animation Control
Interaction with Lipsyncing System
No Interaction with Recorded Video Playback
Final Video Output Generation

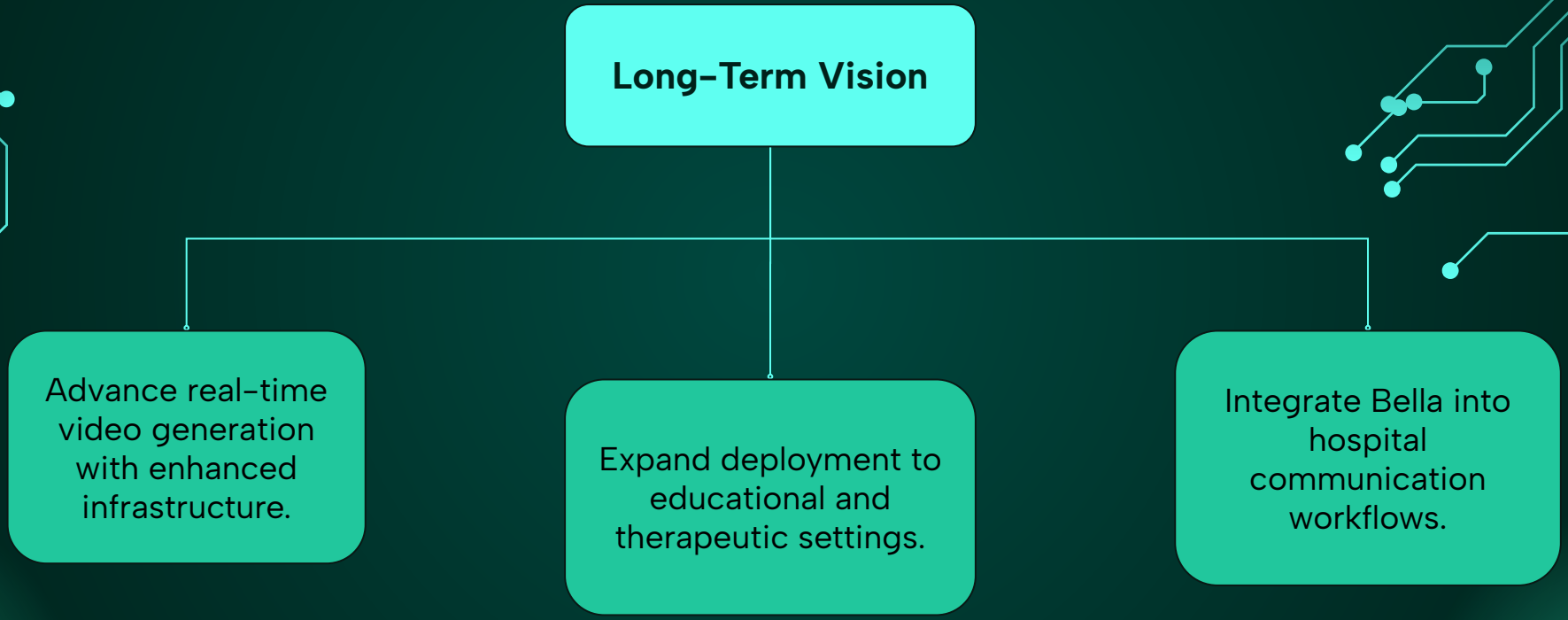
05

FUTURE DEVELOPMENT

FUTURE DEVELOPMENT



FUTURE DEVELOPMENT



06

CONCLUSION

Project Bella showcased the potential of **holographic companions** to bring warmth to hospital environment.

Despite challenges in **real-time video generation**, the **clip-based system** using **CogVideo** and **Wave2Lip** enables engaging, adaptive interactions.

Future improvements will address current limitations and explore broader applications.

Abdellahi El Moustapha

Baptiste LANGLOIS

Remi Uttejitha ALLAM

Likhita YERRA

Brandon Ngahi

Ahmed Aziz Ben Aissa