

INTRODUCTION TO LIPGAN

Significance in Speech Synthesis Technology

LipGAN Researchers

LIPGAN was developed in the Year OCTOBER 2019

The development of LipGAN involved several key researchers and contributors



Zhihong Li



Wen Liu



Gan Chee Lip



Pavlo Molchanov



Jan Kautz

DEFINITION

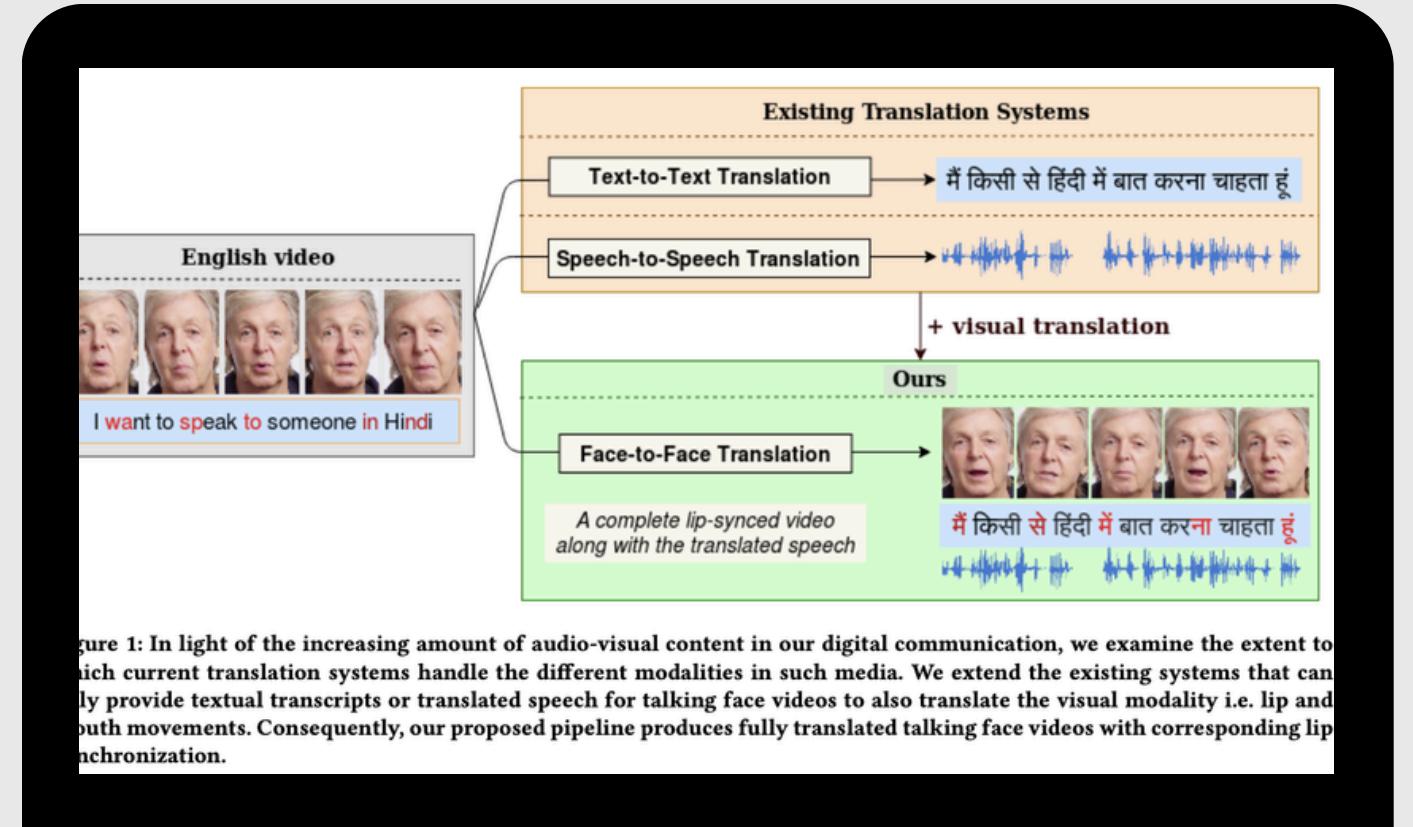
- ⌚ Lip-syncing using generative adversarial networks.



INTRODUCTION:

OBJECTIVE:

- Generate realistic lip-sync GANs.
- Explore potential applications and address ethical concerns.



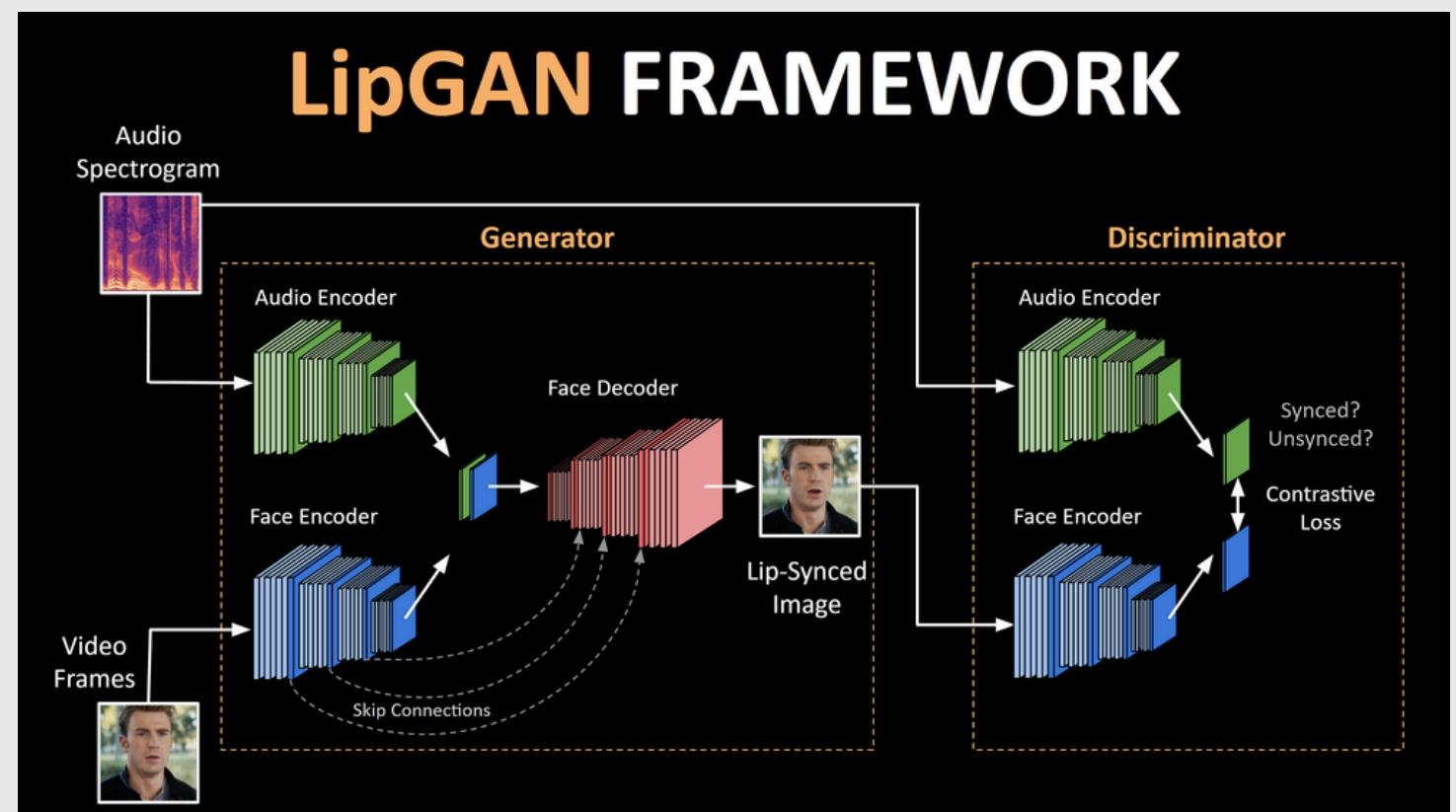
LIPGAN FRAMEWORKS:

Utilization:

- Various frameworks provide tools ---> building, training, and deploying DLM.

Purpose:

- Frameworks enables ---> generation of realistic lip movements synchronized with audio.



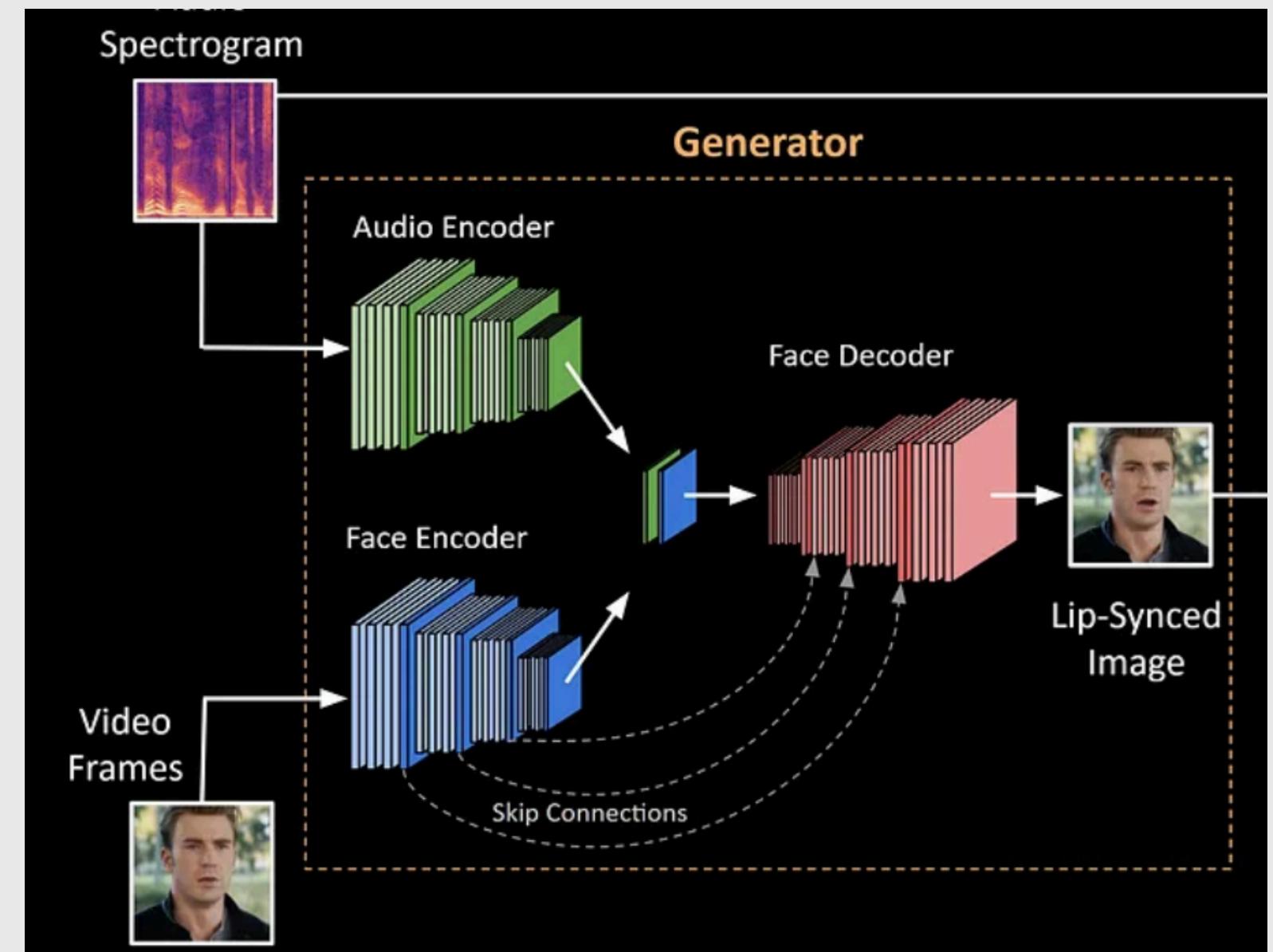
LIPGAN FRAMEWORKS:

GENERATOR:

Generates realistic lip movements synchronized with the input audio.

ARCHITECTURE COMPONENTS :

- **Audio Encoder:** Processes audio input to extract relevant features.
- **Video Encoder:** Processes video input to extract facial features.
- **Fusion Layer:** Combines audio and video features.
- **Lip Movement Generator:** Synthesizes lip movements based on fused features.
- **Output Layer:** Produces the final video frames with synchronized lip movements.



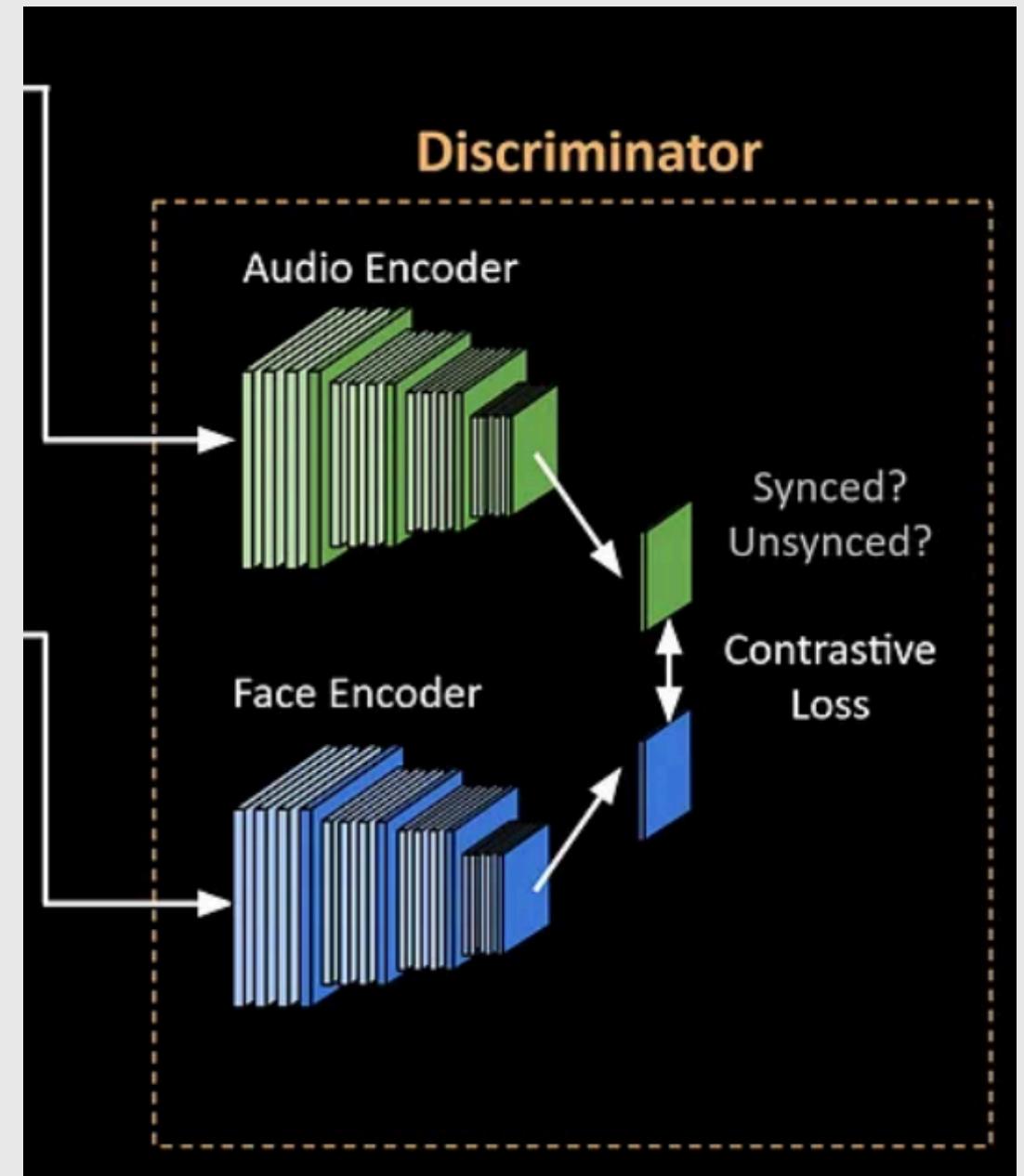
LIPGAN FRAMEWORKS:

DISCRIMINATOR:

Distinguishes between real and generated lip movements.

ARCHITECTURE COMPONENTS :

- **Input Layer:** Receives both real and generated video frames.
- **Feature Extractor:** Extracts features from the input frames.
- **Classification Layer:** Determines the authenticity of the input (real or generated).
- **Loss Function:** Provides feedback to the Generator to improve realism.



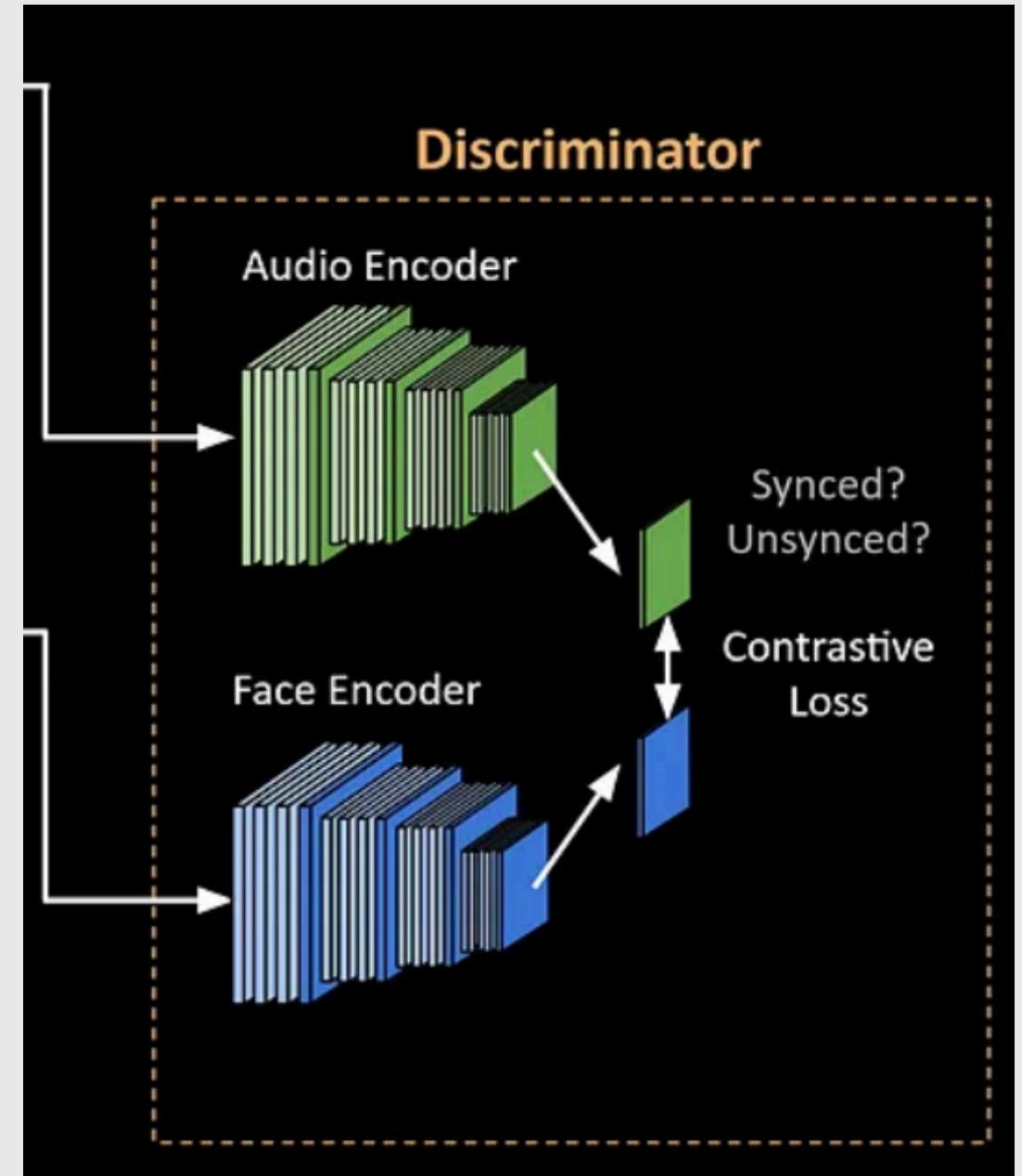
LIPGAN FEASIBILITY

BENEFITS:

Video editing

CHALLENGES AND LIMITATIONS:

- Training data
- Real-time processing
- Accuracy



SAMPLE DATAS

Data Sources:

- Public datasets used (e.g., GRID, LRS2).

Preprocessing Steps:

- Data cleaning, alignment of video frames with audio, extraction of facial landmarks, normalization.

SAMPLE DATAS

Set	Dates	# utterances	# word instances	Vocab
Pre-train	11/2010-06/2016	96,318	2,064,118	41,427
Train	11/2010-06/2016	45,839	329,180	17,660
Validation	06/2016-09/2016	1,082	7,866	1,984
Test	09/2016-03/2017	1,243	6,663	1,698

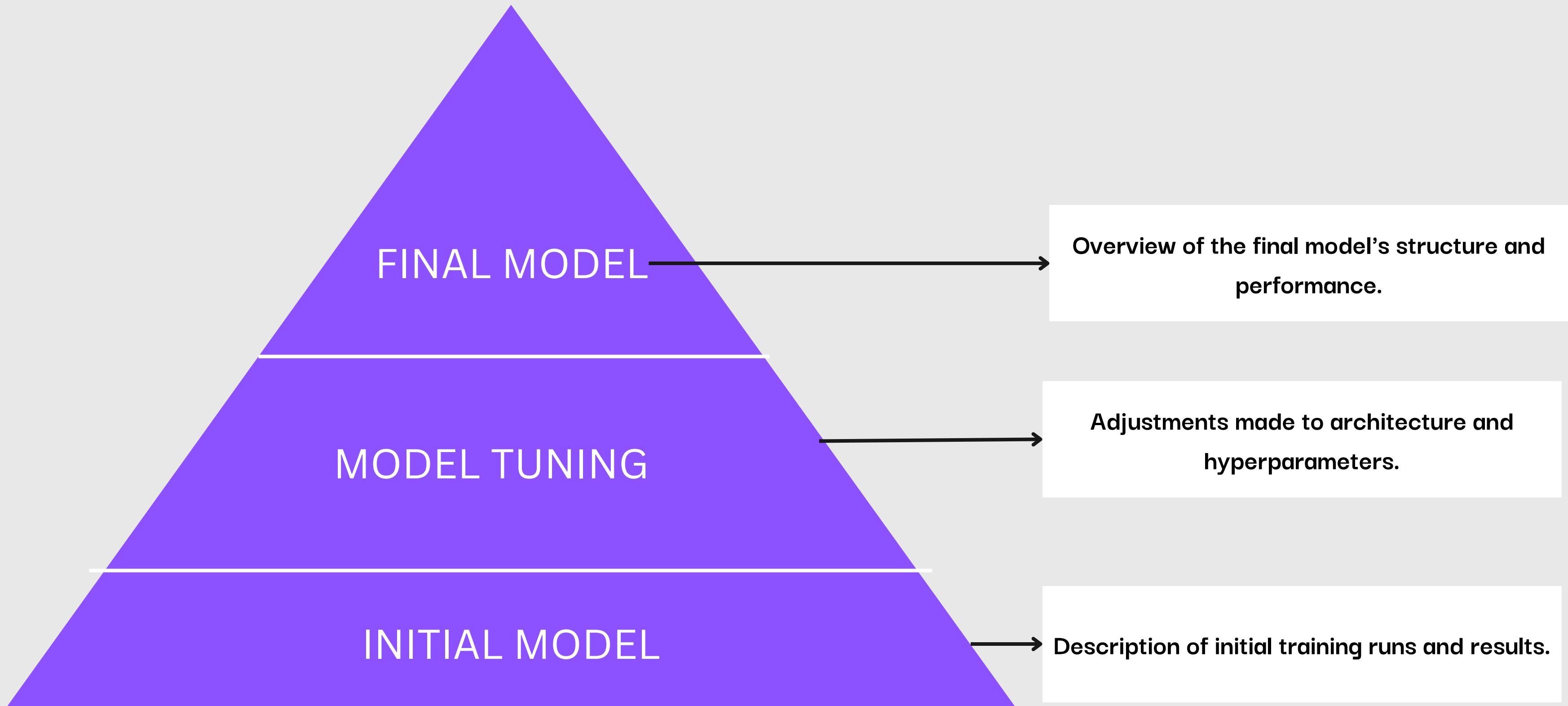
MODEL IMPLEMENTATION

ARCHITECTURE:

- Key components of the LipGAN model.
- Incorporation of GANs, attention mechanisms, and specific layers for lip-sync generation.

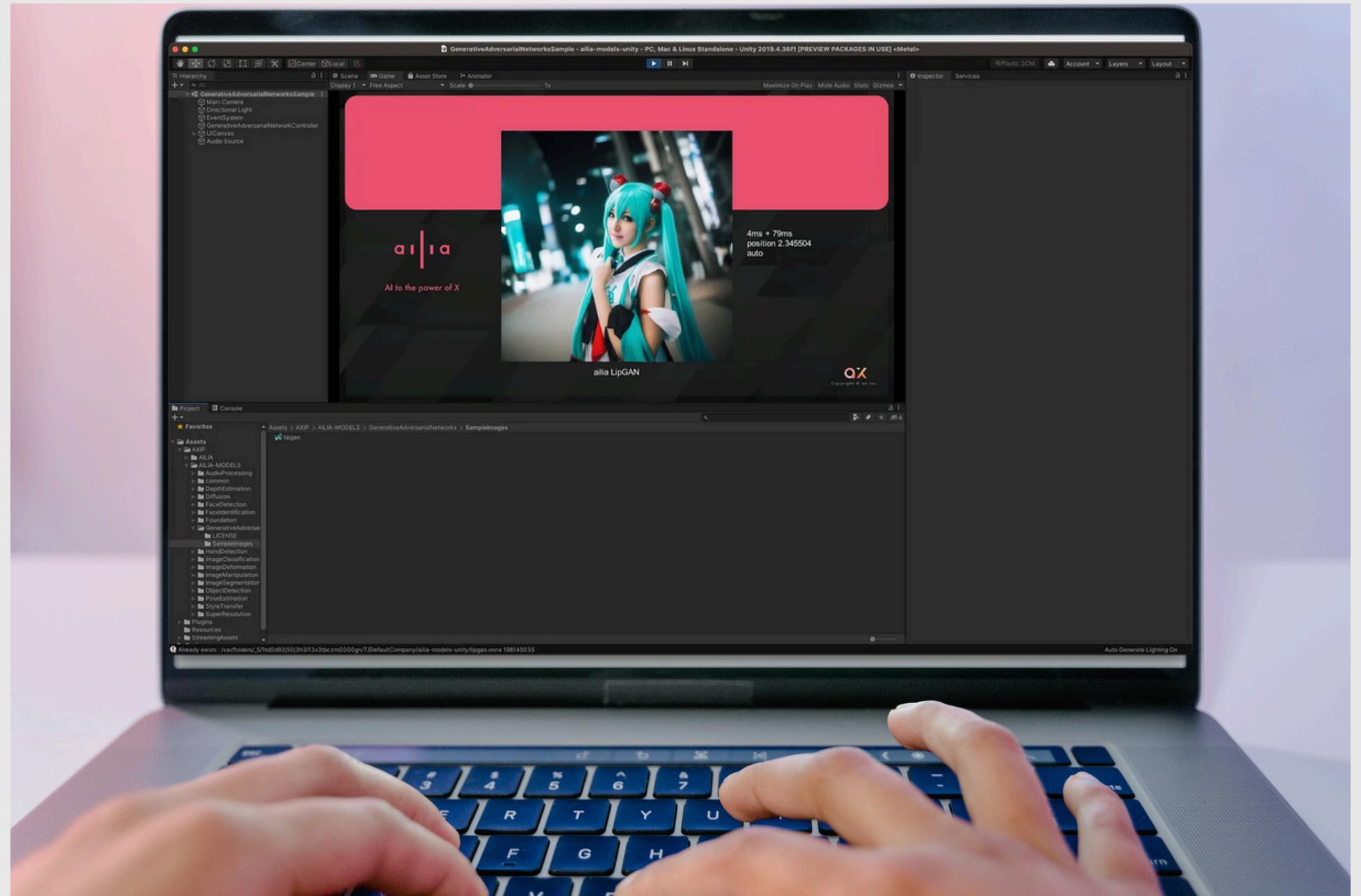


TRAINING PROCESS:



USAGE WITH UNITY

- Unity integrates LipGAN for real-time lip-sync.
- Unity visualizes and tests lip-sync.



RESULTS

Quantitative Results:

- Improvement in metrics from initial to final model.

Qualitative Results:

- Examples of generated lip movements with corresponding audio.

GITHUB LINK

<https://github.com/Staines17/Kaizen-Voiz-LipGan>

THANK YOU

