# VoiceMod

Alejandro Miranda

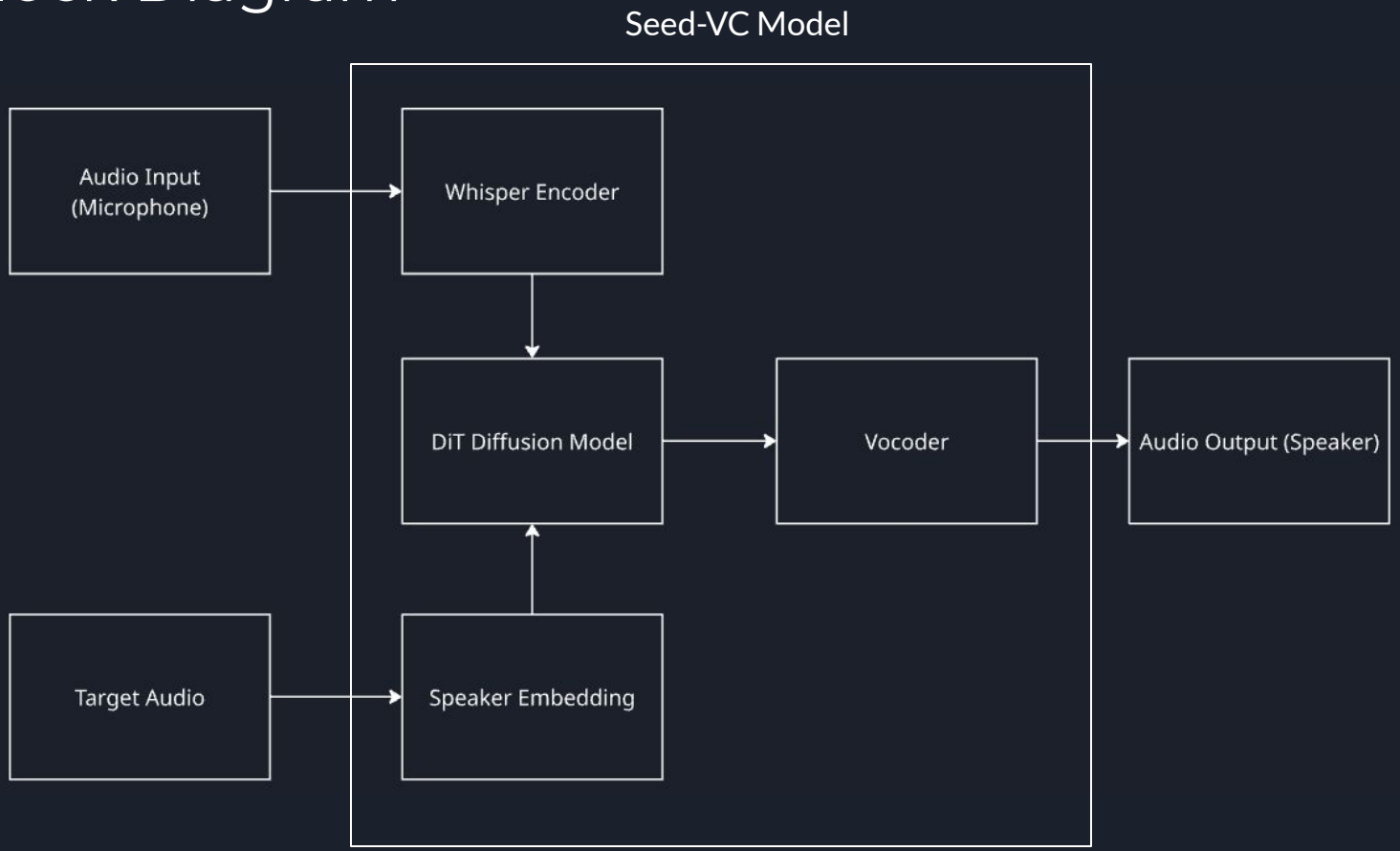
# Abstract

- What: Real-time speech-to-speech voice conversion system running on embedded hardware
- Why: Privacy-preserving, low-latency voice transformation for entertainment purposes
- For Whom: Entertainment/Recreational Users
- Product-type: Proof-of-concept

# BLERP

- **B - Bandwidth:** No network requirement
- **L - Latency:** Real-time requirement
- **E - Economics:** No cloud API keys
- **R - Reliability:** No internet connectivity issues
- **P - Privacy:** All on device

# Block Diagram



# Data Collection

- YouTube + VocalRemover.org + Audacity
- ML Model: Seed-VC
- Embedded ML Implementation: Jetson Orin Nano
  - Optimizations for Model
    - TensorRT Optimization
    - Faster CFM (Heun)
    - Flash Attention for Faster Attention Computation

# Validation Accuracy: 73.2%
# Testing Accuracy: 71.7%

- Compared Original Recordings to Output with **Same Phrasing and Words Said**

Original:

Output:

# Challenges

- Setting up the Jetson Environment to Run the Model
- Optimizing the Speed of the Model
- Balancing Quality vs Latency
- Microphone and Speaker Connection Issues