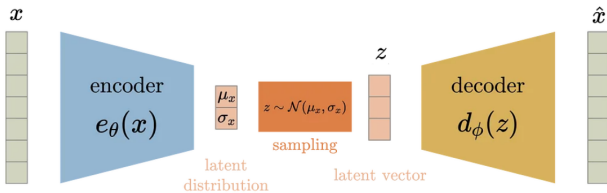


# Vits

Conditional Variational Autoencoder with  
Adversarial Learning for  
End-to-End Text-to-Speech

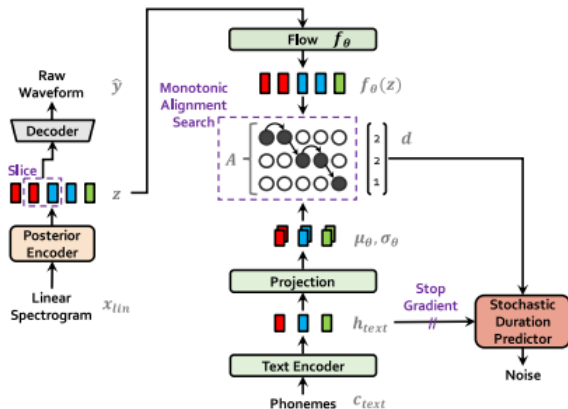
# VAE



# Vits

- ▶ **Posterior Encoder and Normalizing Flow**
- ▶ **Stochastic Duration Predictor**
- ▶ **Variational and Adversarial Training**
- ▶ **High-Fidelity Generator**
- ▶ **Monotonic Alignment Estimation**
- ▶ **Multispeaker Capability**

# Vits



# Processo de Treinamento

## **Primeira Tentativa**

GPU: L40S

Dataset: LJ Speech

60 épocas

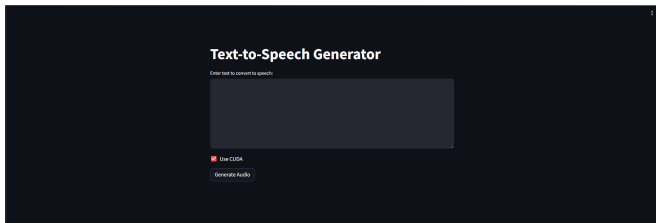
## **Segunda Tentativa**

GPU: L40S

Dataset: CETUC

20 épocas (Problemas com GPU)

# Aplicação



link