

# NVIDIA Riva TTS



Kindly refer the link below:

<https://www.youtube.com/watch?v=2M85UoMzthM>

NVIDIA Riva: Using Speech AI for Transcription, Translation, and Voice  
Highlight Reel Demo Video

# TEXT-TO-SPEECH IS EVERYWHERE!

Hundreds of Billions of Minutes of Speech Generated Daily



## Contact Center

500M Calls Daily



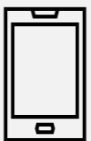
## Consumer Applications

1.8B Daily Minutes



## Digital Avatars

\$10B Market 2020



CRISP / Service Provider



Education



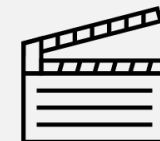
Energy



Finance



Healthcare



Media & Entertainment



Retail

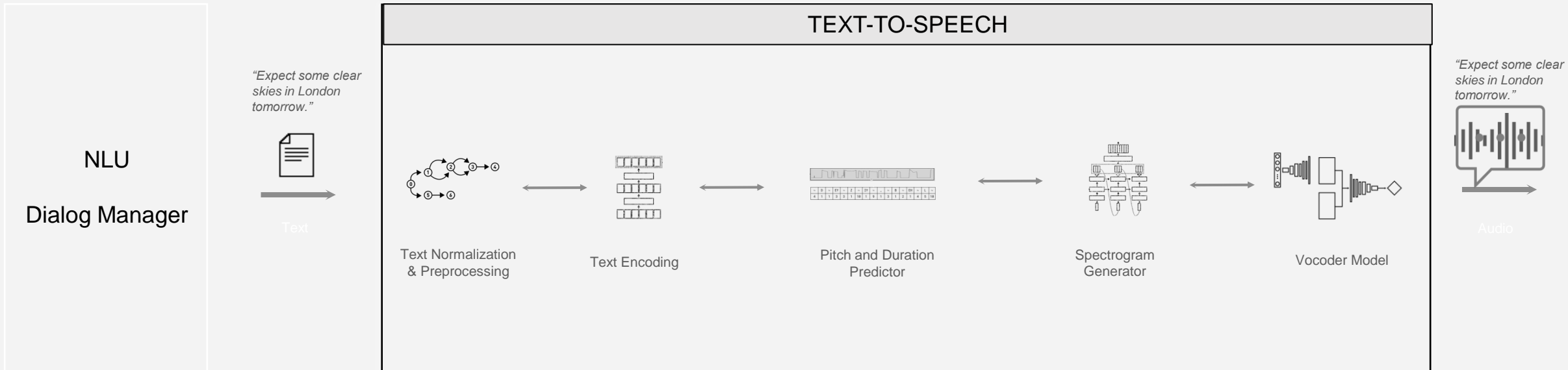


Telecom



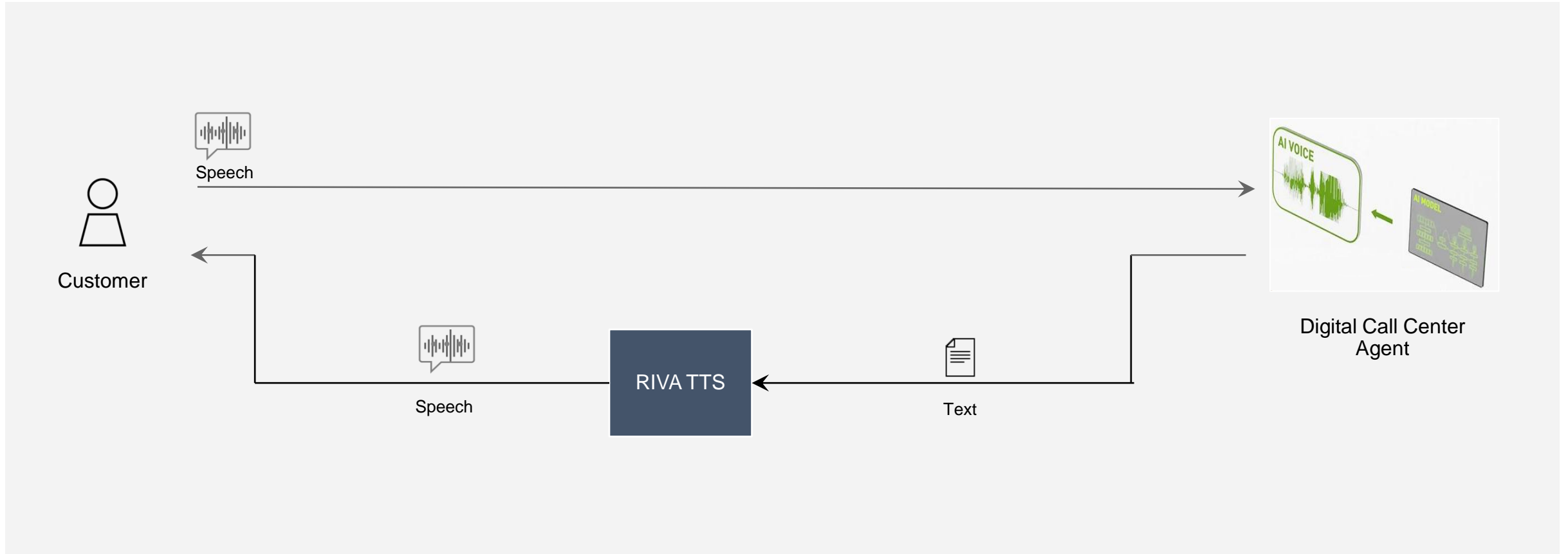
Transportation

Sounding Naturally | Emotion Expressivity | Compute Intensive | Pipeline Robustness

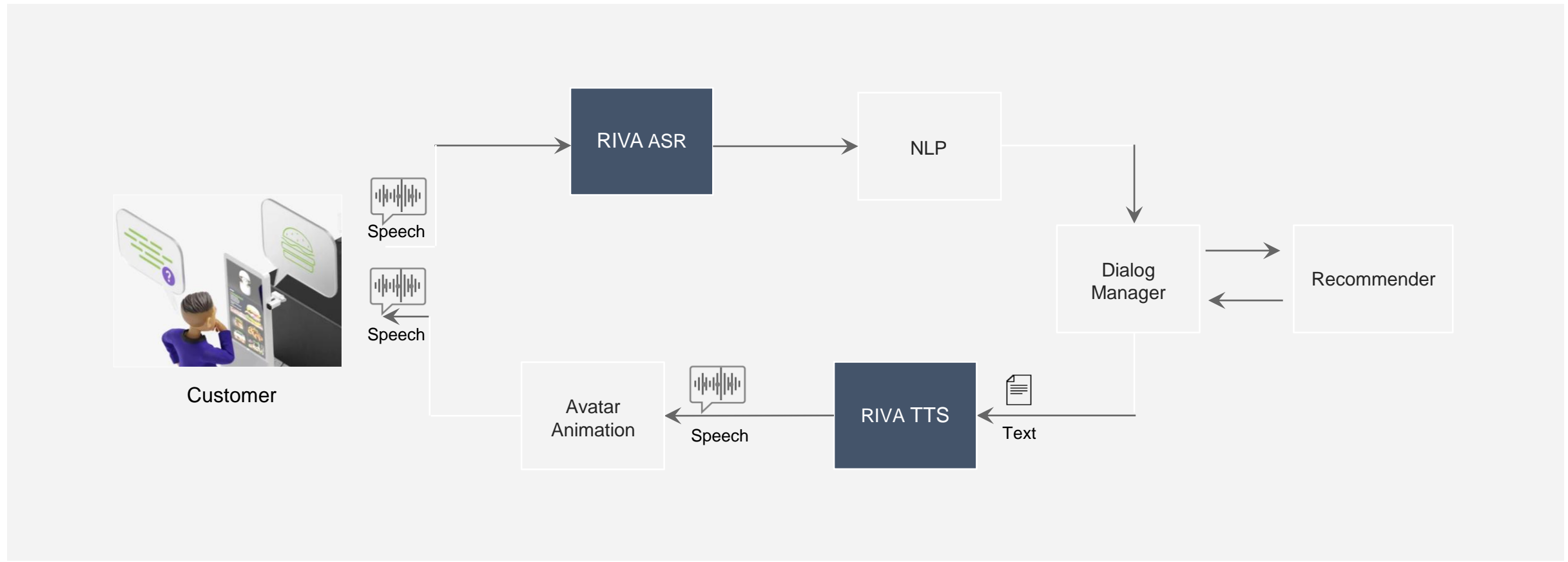


NLU – Natural Language Understanding

## Give Human Voice to Digital Assistants and Chatbots



TTS – Text-To-Speech



ASR – Automatic Speech Recognition | NLP – Natural Language Processing | TTS – Text-To-Speech

## Customer Pain Points



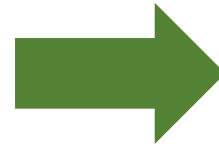
Robotic Voices Due to No Access to SOTA Speech Models



Tight Latency for Natural Conversations



Flexible Scalability & Data Privacy Issues



## World-Class TTS



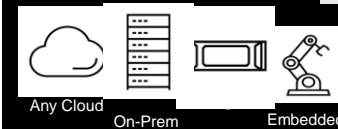
Human-Like Voices with From-Data-To-Model TTS Pipeline Customization



Real-Time Latency Delivery for Natural Conversations



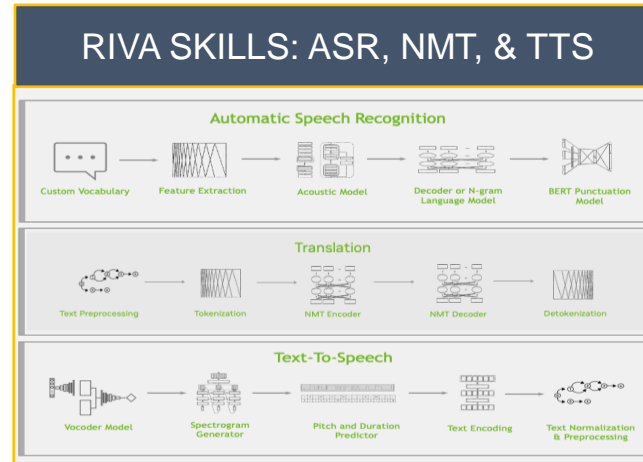
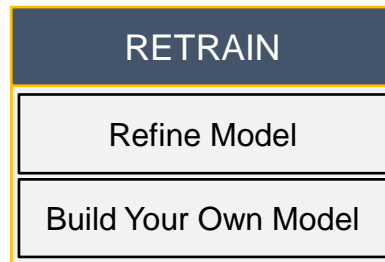
Large-Scale Deployment on Customer-Choice Infrastructure



## Fully Customizable, GPU-Accelerated SDK for Real-Time Speech and Translation AI

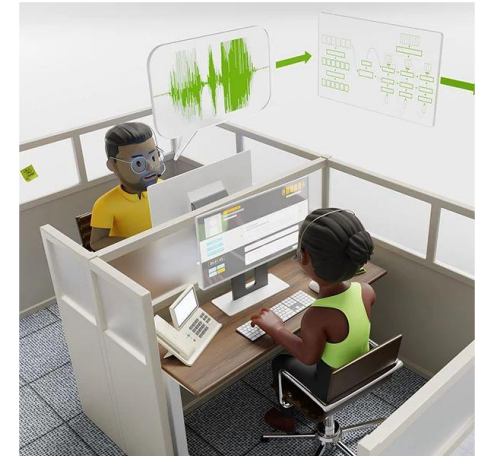


Out-Of-The-Box



Accurate  
Transcriptions &  
Translations

Engaging Voice



Train for Any  
Domain and Language

- Fully Customizable for the Best Possible Accuracy
- GPU-Accelerated, Real-Time
- Scale to Hundreds of Thousands of Users
- On-Prem, in Any Cloud, at the Edge, or Embedded

SOTA – State-Of-The-Art | ASR – Automatic Speech Recognition | NMT – Neural Machine Translation | TTS – Text-To-Speech



# NVIDIA RIVA TEXT-TO-SPEECH

World-Class TTS with SOTA Deep Learning

## Applications

Contact Centers

Consumer Applications

Digital Avatars



CRISP / Service Provider



Education



Energy



Finance



Healthcare



Media & Entertainment

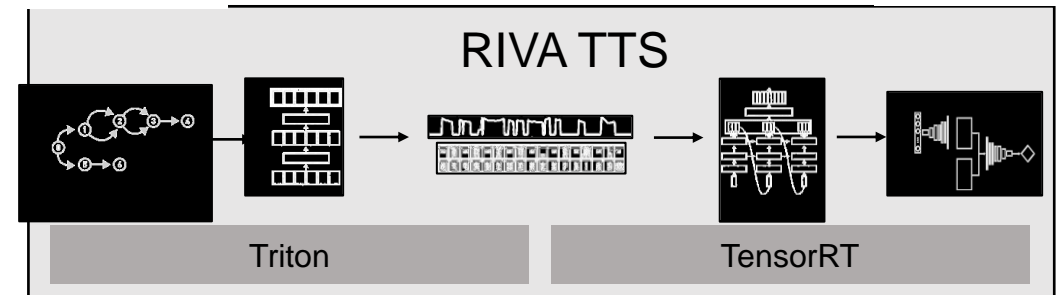


Retail



Transportation

- **SOTA OOTB** expressive voice **models**
- Female & male **OOTB professional voices** for English
- **Easy voice control** at inference time with SSML API for:
  - Pronunciation
  - Pitch
  - Volume
  - Pause
- **12X higher inference performance** versus existing technologies.



TTS – Text-To-Speech | SOTA – State-Of-The-Art | OOTB – Out-Of-The-Box | SSML – Speech Synthesis Markup Language

Consumer Application | Healthcare | UCaaS | Telecommunications



### Local English-Singapore Voice for Driving App

- TTS Voice with Correct Pronunciation, Tone, & Accent
  - 10s of 1000s of Monthly Active Users
  - 1000s Concurrent Users
  - K8s Scalable Cloud Deployment
  - Portability to Any On-Premises or Cloud



### 50% Reduced Patient Waiting Time

- TTS for Check-In Kiosk & Patient Announcements
  - Used by Top 10 America's Best Hospitals as per U.S. News & World Report
  - Coordinating Care of 1K+ Patients/Day
  - Improved Operations & Patient Experience

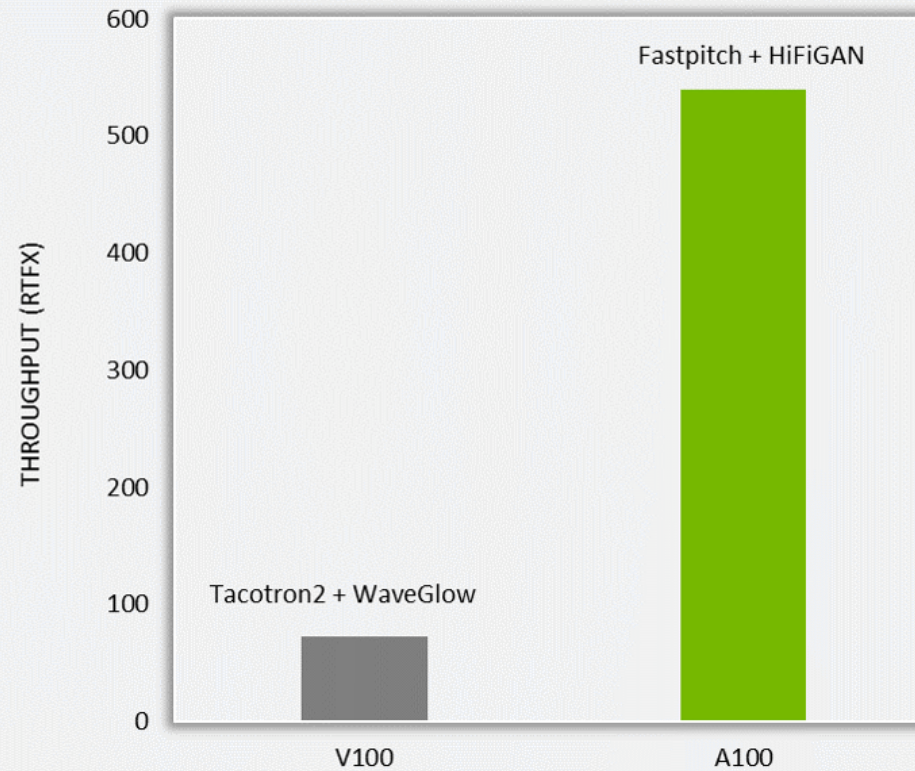


### SOTA WER for Noisy Environments & Expressive Synthetic Voices

- ASR & TTS for Consumer Application
  - 100M Monthly Active Users
  - Integrated in voice-enabling Dev Tool
  - 16 Custom Voices for Engaging Communication

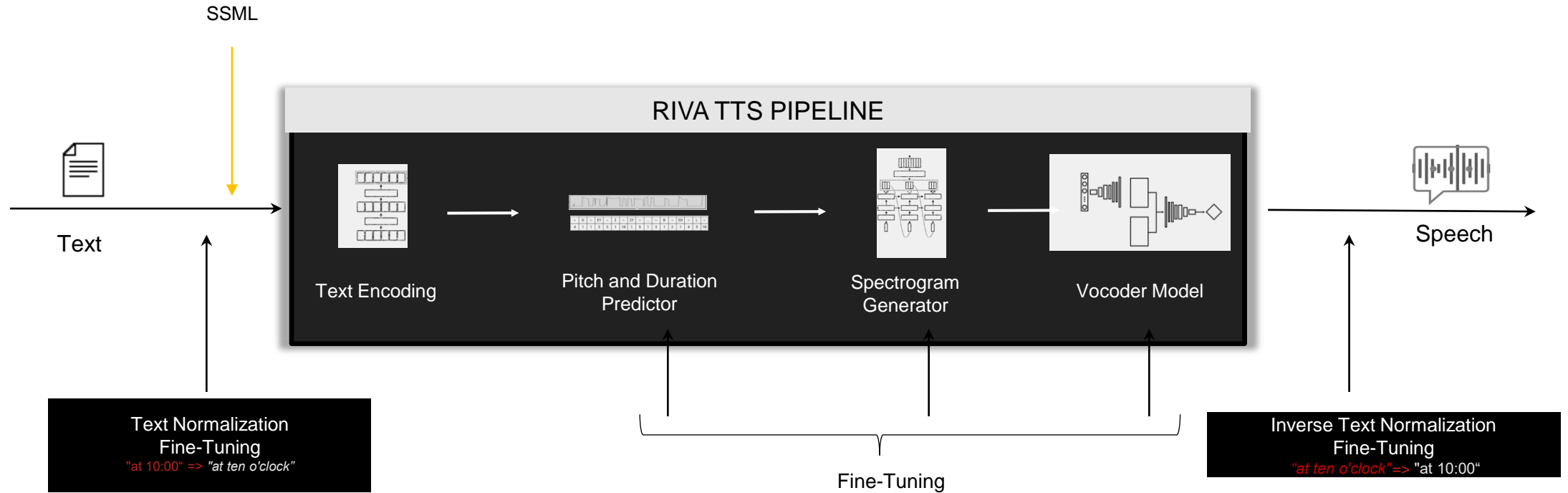


## 12X Performance with SOTA TTS Models





TTS – Text-To-Speech | SOTA – State-Of-The-Art | RTFX – Inverse of Real Time Factor (RTF) =  $\frac{\text{compute time [sec]}}{\text{audio inferred [sec]}}$

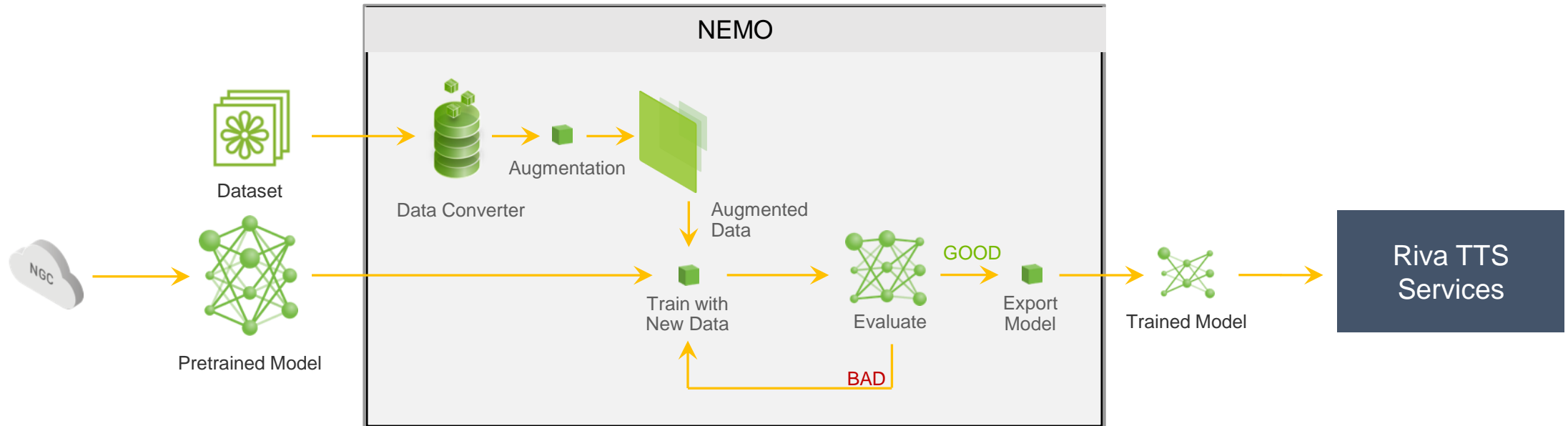
•Controllability | Expressiveness | High-Performance



Customization during:

-  Training (server-side) – Need to rebuild/restart server
-  Inference (client-side) – No need to rebuild/start Riva server

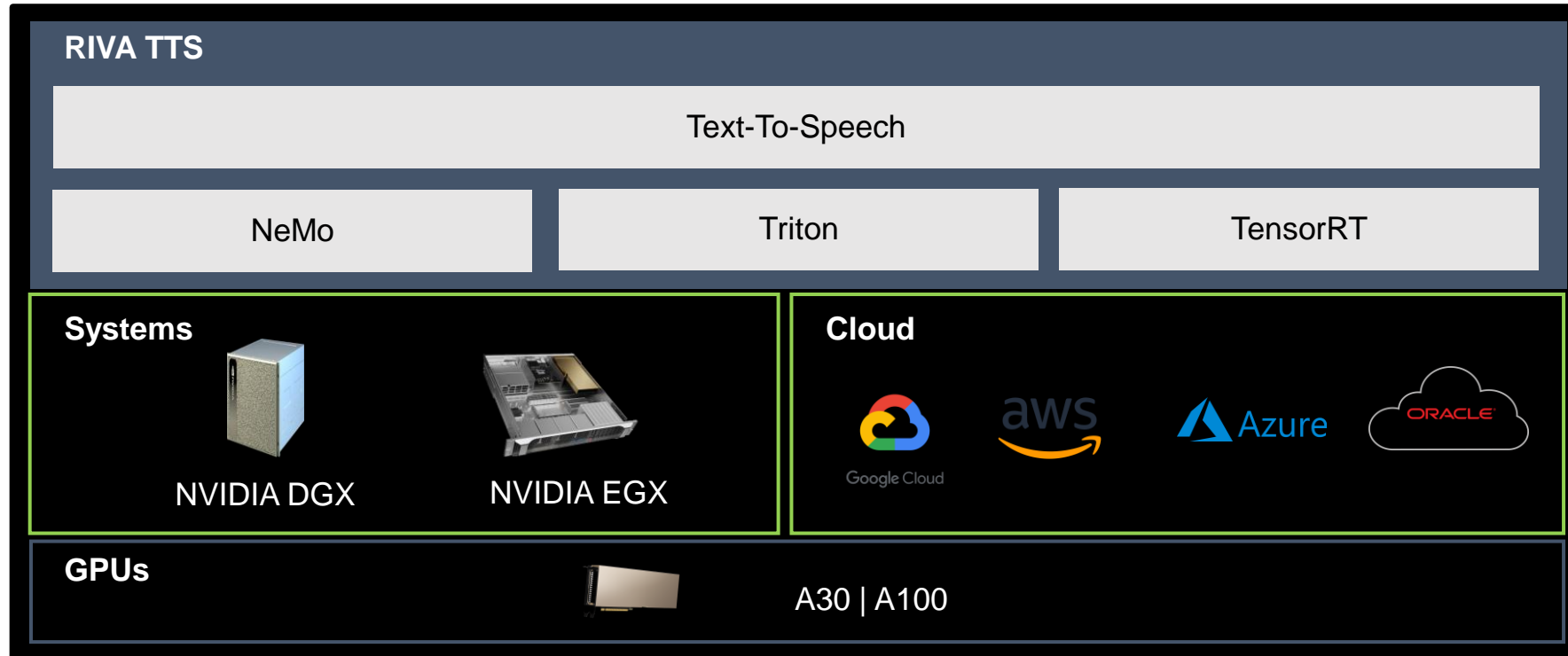
Low-Code | GPU-Optimized | Deploy to Riva



TTS – Text-To-Speech | NGC – NVIDIA GPU Cloud

# RIVA TTS RUNS ON ANY CLOUD, ON-PREMISES, AT THE EDGE, AND EMBEDDED

GPU-Optimized for Minimum Latency and Maximum Throughput

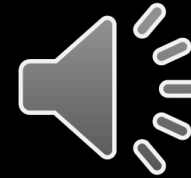


## CUSTOMIZABLE TTS

Kindly refer the link below:

<https://www.youtube.com/watch?v=bt8ZwjuFOmQ>

## HIGH-QUALITY TTS VOICES



US English Female



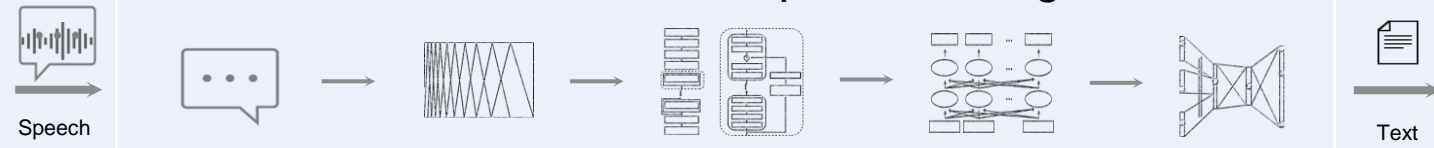
US English Male

<https://www.nvidia.com/en-us/ai-data-science/products/riva#demos>

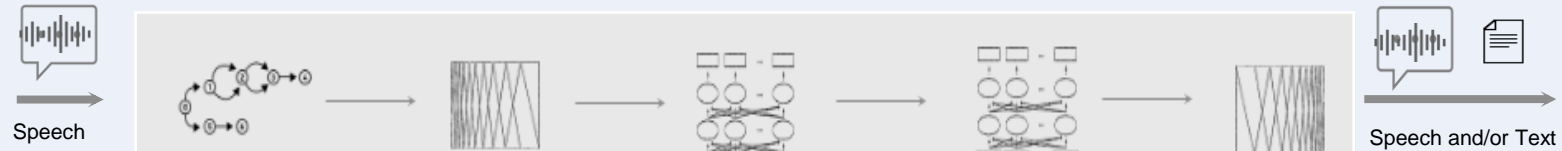
# NVIDIA RIVA SOFTWARE

Enterprise-Grade Security, Reliability, Stability and Support

## Automatic Speech Recognition



## Neural Machine Translation



## Text-To-Speech



Unlimited Usage



Scale to any cloud & on-premises

Access to NVIDIA AI Experts



8-5 local business hours  
guidance on configurations,  
performance

Control Maintenance &  
Upgrade Schedule



Long Term Support

Ticket Prioritization



Latest security fixes, and  
maintenance releases



---

# THANK YOU!

## EASY SPEECH AI CUSTOMIZATION FOR LOCAL SINGAPOREAN VOICE

Creating a driving app with a local experience for Singaporean drivers is challenging since Singapore's flavor of English consists of a mix of words from English, Malay, Hokkien, Mandarin, and Tamil.

NCS, a multinational company and a technology partner of the Singapore government for various transpiration projects, designed Breeze, a local driver's app with a native Singaporean voice.

NCS customized the NVIDIA Riva FastPitch model with prosody control over pitch and phoneme duration and built its own TTS engine for English-Singapore using local speakers' voice data. Breeze runs with Riva-scalable Kubernetes on AWS EKS with GPUs and is portable to any containerized on-premises or cloud environment.

Today, tens of thousands of monthly active users—with thousands accessing concurrently—enjoy Breeze with the correct Singaporean pronunciation of local phrases and street and road names.

