HIGHLY ACCURATE, REAL-TIME SPEECH AI ENHANCES CALL CENTERS

"We had great support from the NVIDIA team while fine-tuning our ASR model on top of NVIDIA SDKs. We reduced the word error rate (WER) and increased concurrency. Additionally, we reduced response time in Singaporean English by 38%—from 260ms to 162ms"

— Jimmy Padia, CEO and Founder, Floatbot.AI

Using Speech AI to Deliver Exceptional Customer Support

Floatbot, a leading unified voicebot and chatbot platform, specializing in conversational AI, drives growth for insurance and fintech clients by enhancing their customers' experiences. To provide delightful customer experiences, automated call centers often need to improve their automatic speech recognition (ASR) accuracy and real-time response—Floatbot sought to do both.

NVIDIA Solution

To boost ASR accuracy and server concurrency, and deliver real-time results, Floatbot began using pre-trained models for their customers in Singapore. In addition to the acquired National Speed Corpus (NSC) data set, Floatbot collects more data through campaigns, which its research team fine-tunes using the NVIDIA TAO AI-model-adaptation framework. Floatbot also uses the NVIDIA® Riva Speech AI SDK to build customized speech applications that deliver real-time performance.

Floatbot Results

Floatbot experienced a 30% improvement in accuracy over a large cloud-based ASR service provider for Singlish using Riva pre-trained models. Latency, which included network traffic and inference, was reduced from ~260ms with the cloud-based service to ~162ms with Riva.

About Floatbot

Floatbot is a leading unified voicebot and chatbot platform with speech AI capabilities. Their SaaS-based, no-code speechAI platform helps insurance carriers and fintech across the globe increase digital sales, automate inbound and outbound calls, and deliver delightful customer support.

Floatbot improves ASR accuracy and the response time of its unified voicebot and chatbot platform.

INDUSTRY

> Insurance, Fintech

CHALLENGE

> Improve ASR accuracy and real-time response for automated call centers

PRODUCTS USED

> NVIDIA V100 Tensor Core GPU
> NVIDIA Riva SDK
> NVIDIA Nemo SDK
> NVIDIA TAO Toolkit
> PyTorch
> TensorFlow

RESULTS

> 38% faster ASR results
> 30% improvement in accuracy

Learn more

developer.nvidia.com/riva