NVIDIA

SPEECH AI FOR AWARD WINNING CUSTOMER CARE

"With NVIDIA Riva services, fine-tuned using T-Mobile data, we're building products to help us resolve customer issues in real time. After evaluating several automatic speech recognition solutions, T-Mobile has found Riva to deliver a quality model at extremely low latency, enabling experiences our customers love."

Matthew Davis, Vice President of Product and Technology, T-Mobile

Assisting Agents in Real-Time

In millions of calls with global customers daily, consumers expect agents to fix their issues quickly and efficiently. Enterprise leaders are looking into automated and scalable solutions to assist customer care agents in providing the best experience.

T-Mobile offers award-winning customer service in Customer Experience Centers and aims never to have the customer repeat themself. To empower their frontline agents, T-Mobile implemented Expert Assist. This AI-based software transcribes real-time customer conversations and recommends solutions to thousands working on the frontline.

NVIDIA Solution

T-Mobile uses NVIDIA Riva, the GPU-accelerated SDK for building customized speech applications, and NVIDIA NeMo, the open-source framework for building, training, and fine-tuning state-of-the-art conversational AI models. These NVIDIA tools enabled T-Mobile engineers to fine-tune automatic speech recognition models on their custom datasets and interpret customer jargon accurately across noisy environments.

T-Mobile Results

With NVIDIA NeMo, the T-Mobile team easily retrained speech AI models and achieved unparalleled accuracy across noisy production environments and jargon, slashing the word error rate (WER) by 10%. The overall latency during inference is reduced by 10X using NVIDIA Riva, enabling the highest level of performance of more than 10,000 concurrent streams in real-time.

About T-Mobile

T-Mobile is a supercharged Un-carrier, delivering an advanced 4G LTE and transformative 5G network. T-Mobile offers industry leading customer service delivering the best possible service experience with an unmatched combination of value and quality.

T Mobile

T-Mobile enables customer care agents to solve queries quickly with recommendations based on highly accurate NVIDIA Riva ASR-generated transcripts.

INDUSTRY

> Telecommunications

CHALLENGE

> **Expert Assist** needed more accurate transcriptions in order to offer effective recommendations to customer care agents.

PRODUCTS USED

> NVIDIA Riva SDK
> NVIDIA NeMo SDK
> NVIDIA A100 Tensor Core GPU

RESULTS

> 10% higher accuracy
> 10X overall latency improvement

Learn more

https://developer.nvidia.com/riva