IOWN INTEGRAL

NTT R&D FORUM 2024

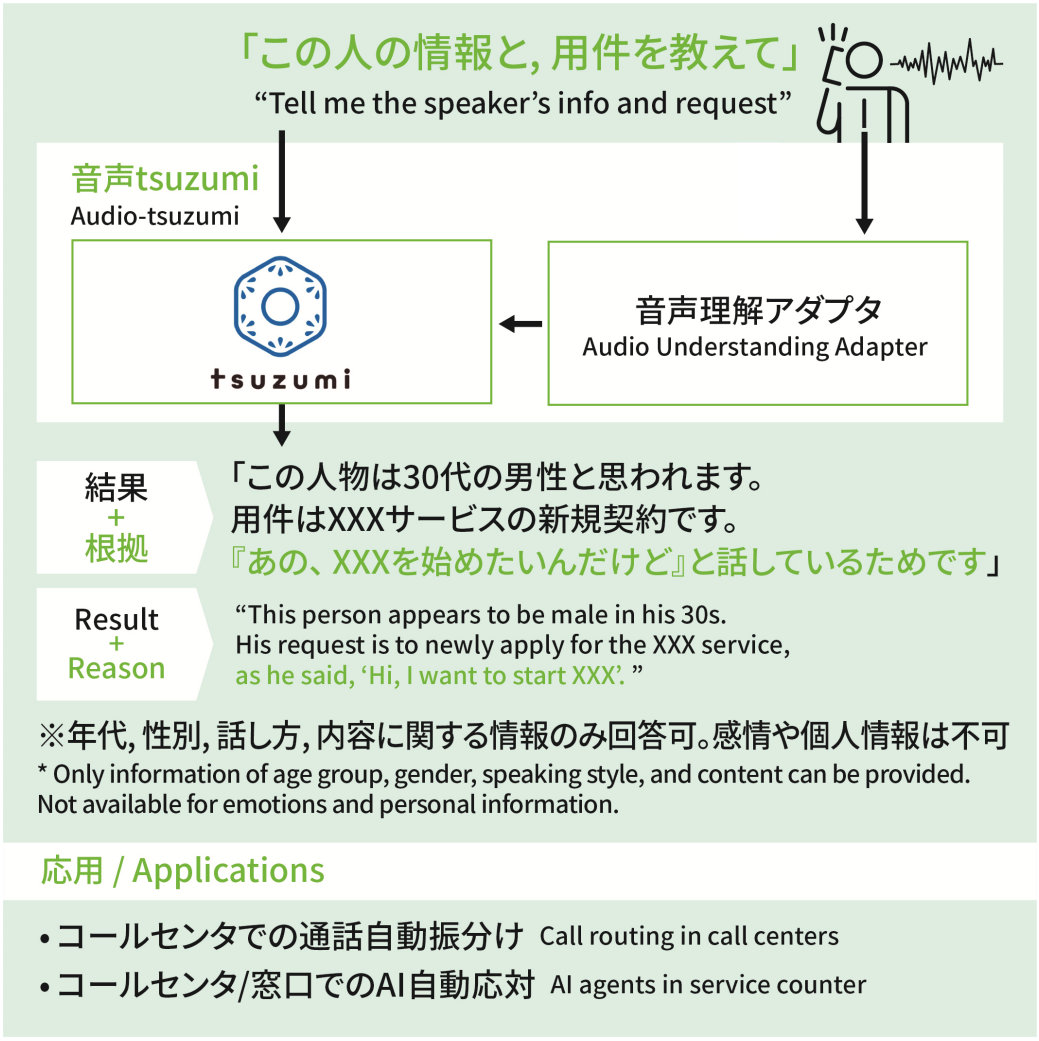
DEVELOPMENT

δ 01-12

Audio-tsuzumi: Speech understanding LLM

Understand a wide range of speech information to respond like a native Japanese speaker

#Productivity Improvement #Customer Experience Value Creation



///Technical Issue

Conventional speech processing methods provide individual predictions of speech information, making it difficult to generate integrated predictions and/or reasoning.

///Research Goal

Create a new business, such as AI call routing (market: $1.6B), using LLM-based speech understanding technology that can respond with reasoning/evidence.

---Technology

1. A novel speech understanding model composed of a speech encoder, an LLM, and a bridge network.
2. Open-QA data generation methods for training.
3. Training strategy to learn relationship between speech features and textual representation.

---Novelty

A novel speech understanding model that has been accepted at top-tier international conferences and generating training data specialized for the Japanese language.

---Applicable Business

In the call center field, we realize AI call routing that directs calls based on natural spoken language. Compared to traditional IVR, this reduces the burden on users and decreases routing errors (Global AI call center market: $1.6B in 2023).
In the future, we aim to develop AI agents and AI customer services that can respond through voice interaction.

Exhibitors= NIPPON TELEGRAPH AND TELEPHONE CORPORATION

Contact URL