



# CARTGPT

TECHNOLOGY NUMBER: 2025-161



## Technology ID

2025-161

## Category

Software  
Software & Content  
Accessible Technologies/Sound  
Accessibility

## Inventor

Dhruv Jain  
Liang-Yuan Wu

## Further information

Ashwathi Iyer  
[ashwathi@umich.edu](mailto:ashwathi@umich.edu)

## OVERVIEW

A real-time captioning enhancement system, CARTGPT, uses AI to correct errors and fill gaps in human-generated transcripts for deaf and hard of hearing (DHH) users.

- Integrates outputs from human captioners and speech recognition, using large language models to improve word accuracy on technical or noisy speech by up to 5–17% over standalone solutions.
- Enables clearer, context-aware captions in demanding environments, creating new opportunities in education, healthcare, conferences, and any setting requiring accessible live communication.

## BACKGROUND

Captioning is a cornerstone accessibility tool for millions of deaf and hard of hearing people worldwide, critical for participation in workplaces, classrooms, hospitals, and public events. Traditionally, professional captioners using Communication Access Realtime Translation (CART) deliver highly accurate live transcripts, often preferred over automated systems due to their ability to reflect speaker cues and capture complex conversations.

However, in real settings—like technical meetings or noisy environments—CART accuracy drops. Captioners struggle with unfamiliar vocabulary, rapid speech, unclear audio, and limited preparation, leading to missing words and reduced comprehension.

Current alternatives—Automatic Speech Recognition (ASR) and crowd-sourced editing—do not match professional accuracy, particularly for technical content. Market demand for better

## [View online](#)



accessibility is growing, driven by disability inclusion laws, online learning expansion, and remote work, with surveys showing persistent gaps in real-time comprehension despite available captioning technology.

## **INNOVATION**

CARTGPT works by continuously receiving two live transcripts: one from a human captioner, and another from ASR software. When mistakes or gaps are detected in the human transcript, CARTGPT uses a large language model (LLM)—the same technology behind advanced chatbots—to cross-reference both transcripts and surrounding conversation, instantly generating plausible corrections or filling missing information.

Unlike previous solutions, CARTGPT harnesses the strengths of both expert humans (contextual, speaker-aware) and AI (broad vocabulary, context learning), resulting in substantially improved caption accuracy—even in technical domains and challenging audio. In tests, CARTGPT outperformed both traditional CART and state-of-the-art ASR in word accuracy and user comprehension, doing so with almost imperceptible delay. This innovation enables reliable, real-time accessibility for DHH users, with potential for user personalization, transparency, and effortless integration into existing workflows.

## **ADDITIONAL INFORMATION**

### **REFERENCES:**

["CARTGPT: Improving CART Captioning using Large Language Models"](#)

["CARTGPT: Real-Time Correction of CART Captions Using Large Language Models"](#)

### **INTELLECTUAL PROPERTY:**

Patent application pending.