

CapGenie: Realtime Caption Generator

**Prof. Amarja Adgaonkar¹, Shazmeen Shaikh², Manu Kumar Mishra³,
Viraj Bhor⁴, Mohd Amin Shaikh⁵**

¹Professor, Department of Information Technology

²³⁴⁵Bachelor of Engineering in Information Technology

K. C. College of Engineering and Management Studies and Research, Thane, Maharashtra, INDIA

Abstract: *The increasing reliance on multimedia content has heightened the need for inclusive technologies, particularly for individuals with hearing impairments. This paper presents a Real-Time Caption Generator that leverages a Whisper ASR model to convert video audio into accurate, synchronized captions with <500ms end-to-end latency. The system addresses accessibility gaps through a React-based interactive interface, enabling users to customize caption display settings (font style, size, and opacity) for optimal readability. Experimental results demonstrate >90% word accuracy on clean speech datasets.*

Keywords: Automatic Speech Recognition (ASR), real-time captioning, accessibility, hearing impairments, React.js

