

Technical Architecture and Pedagogical Paradigm of AI - Driven English Speaking Platforms

Ved Kumari

Delhi Skill & Entrepreneurship University, India

Abstract: Artificial Intelligence (AI) is transforming Computer - Assisted Language Learning (CALL) by integrating Automatic Speech Recognition (ASR), Natural Language Processing (NLP), neural text - to - speech (TTS), large language models (LLMs), and automated speech assessment into increasingly interactive English - speaking environments. This paper examines the technical architecture and pedagogical paradigm of contemporary AI - driven English - speaking platforms, with particular emphasis on the transition from conventional speech - to - text systems toward specialized pronunciation assessment and conversational AI architectures. The analysis distinguishes semantic - oriented ASR systems from pronunciation - focused engines that employ acoustic feature extraction, forced alignment, phoneme - level modelling, Goodness of Pronunciation (GOP), and Phoneme Error Rate (PER) to provide fine - grained corrective feedback. The study further develops a technical taxonomy of AI English - speaking platforms encompassing phonetic and accent optimization engines, conversational AI tutors and avatar - based systems, professional and assessment simulators, and multimodal general - purpose LLMs. From a pedagogical perspective, the paper examines the potential of AI - mediated interaction to increase speaking practice, reduce foreign - language speaking anxiety, support personalized feedback, and shift pronunciation pedagogy from accent elimination toward communicative intelligibility. At the same time, the analysis identifies important limitations related to acoustic variability, algorithmic bias, pragmatic and cultural competence, computational requirements, privacy, and unequal access to high - granularity AI feedback. The paper proposes a complementary human - AI CALL framework in which specialized AI systems handle repetitive pronunciation, grammar, and fluency practice while human instructors focus on pragmatic, cultural, interpersonal, and higher - order communicative competencies. The study concludes that effective AI - driven speaking education requires an integrated architecture combining technically robust speech assessment, conversational intelligence, pedagogically informed feedback, diverse speech data, and responsible human oversight.

Keywords: Artificial Intelligence, Computer - Assisted Language Learning, AI - Driven English Speaking, Automatic Speech Recognition, Automatic Pronunciation Assessment, Conversational AI, Goodness of Pronunciation; EFL Learning; Intelligibility; Generative AI