# 2. NVIDIA Riva: 최종 대화형 AI 가상 비서 파이프라인 구축

## 프로젝트 개요

NVIDIA Riva Speech AI 서비스를 활용하여 **ASR (자동 음성 인식)**, **Dialog Manager (대화 관리/QA)**, **TTS (텍스트 음성 변환)** 기능을 통합한 엔드투엔드(End-to-End) 대화형 AI 파이프라인을 구축했습니다. 음성 입력에 대해 정확한 질문 응답과 음성 출력을 제공하는 가상 비서를 구현했습니다. (노트북: assessment\_final\_pipeline.ipynb 기반)

## ⚙️ 사용 기술 및 역할

* **주요 기술:** NVIDIA Riva Speech AI (ASRService, TTSService, DialogManager), Python SDK, BERT QA Model
* **핵심 역할:**
  1. **ASR:** 사용자의 음성 파일을 텍스트로 정확히 변환 (Automatic Speech Recognition).
  2. **Dialog Manager (QA):** 변환된 질문과 주어진 문맥(context)을 바탕으로 **BERT 기반 QA 모델**을 사용하여 답변을 추출.
  3. **TTS:** 최종 텍스트 답변을 다시 자연스러운 음성으로 변환하여 사용자에게 출력 (Text-to-Speech).

## 🛠️ 파이프라인 구성 및 흐름

1. **[Input: Audio Sample]** (녹음된 음성 파일)
2. **ASR Predict:** 음성 -> 텍스트 전사(Transcript)
3. **Dialog Manager (QA) Predict:** 텍스트 질문 + 지식 문맥 -> 최종 텍스트 답변 (Extractive QA)
4. **TTS Predict:** 텍스트 답변 -> 합성된 오디오 데이터
5. **[Output: Audio Playback]** (합성된 음성 재생)

## 💡 핵심 성과 및 역량

1. **다중 AI 서비스 통합:** ASR, QA, TTS의 세 가지 독립적인 AI 컴포넌트를 논리적으로 연결하여 연속적인 대화 경험을 구현했습니다.
2. **음성-텍스트-음성 변환 사이클:** 음성 명령을 이해하고, 지식 기반으로 답변을 찾고, 그 답변을 다시 음성으로 전달하는 완전한 대화형 루프를 성공적으로 구축했습니다.
3. **GPU 가속 활용:** NVIDIA Riva의 최적화된 모델을 활용하여 실시간에 가까운 응답 속도를 달성했습니다.

## 📌 주요 코드 로직 (추상화)

# ASR, DM, TTS 클라이언트 정의 및 초기화  
# ...  
  
def run\_pipeline(SAMPLE, context):  
 # 1. ASR (음성 -> 텍스트)  
 transcript, asr\_config = asr\_predict(SAMPLE)  
  
 # 2. Dialog Manager (텍스트 질문 + 문맥 -> 텍스트 답변)  
 dm\_response = dm\_predict(transcript, context[0])  
 dm\_answer = dm\_response.results[0].answer if dm\_response.results else "Fallback Answer."  
   
 # 3. TTS (텍스트 답변 -> 음성)  
 synth\_audio, processed\_text, req = tts\_predict(dm\_answer)  
   
 # 결과 재생 및 저장  
 # ipd.display(ipd.Audio(synth\_audio, rate=sample\_rate\_hz))  
   
 return dm\_answer, processed\_text