

A mesterséges intelligencia fejlődésével egyre elterjedtebbek azok a rendszerek, amelyek természetes nyelvű interakciókat tesznek lehetővé a felhasználó és a gép között. Az ilyen rendszerek egyik modern megközelítése a retrieval-augmented generation (RAG), amely a generatív nyelvi modellek tudását külső forrásból származó információval egészíti ki. A hangvezérlés révén ezek a rendszerek multimodálissá válnak, hiszen képesek a beszéd felismerésére, szöveges feldolgozására és a válasz hangos megjelenítésére is.

A dolgozat célja egy hangvezérelt RAG-rendszer megtervezése és megvalósítása, amely a felhasználó hangbemenetét értelmezi, releváns információkat keres külső tudásbázisokból, majd az eredményt természetes hangú válaszként jeleníti meg.

A hallgató feladatának a következőkre kell kiterjednie:

- Mutassa be a RAG rendszerek elméleti háttérét, architektúráját és alkalmazási területeit.
- Ismertesse a beszédfelismerés (ASR) és a beszédszintézis (TTS) technológiák működését, valamint integrációs lehetőségeiket nyelvi modellekkel.
- Tervezzon és implementáljon egy hangvezérelt információ-visszakereső és -generáló rendszert, amely képes a felhasználói beszédbevitel szöveges feldolgozására, releváns dokumentumok visszakeresésére, és a generált válasz hangos megjelenítésére.
- Elemesse a generált válasz minőségét különböző query-kezelési és retriever-beállítások alkalmazásával.
- Értékelje a rendszer teljesítményét pontosság, válaszidő, és természetes hangvisszaadás szempontjából.
- Foglalja össze a fejlesztés tapasztalatait és tegyen javaslatot a rendszer jövőbeli bővítési irányaira.