

vissza, a koronavírus-járvány első évében került a figyelem középpontjába: úgy bukkant fel, mint a jelentett halálozásnál jobb – ezen belül is különösen: az országok között robusztusan összehasonlítható – mutatója a járvány okozta halálozásnak. Ennek van racionális alapja: a többlethalálozás független a tesztelési intenzitástól és független a haláloki besorolástól, amely problémák valóban megjelennek a regisztrált halálozásban, és valóban nagyban nehezítik az országok közötti összehasonlítását. A dolog azonban nem jön ingyen: a többlethalálozásnak magának is vannak nehézségei, melyek nem kaptak mindig kellő figyelmet. Ezek egy része értelmezésbeli (a többlethalálozás bruttó jellegű mutató, mely egybeméri a járvány direkt hatását, és a járvány, valamint kezelésének indirekt hatásait, ezek miatt elvileg sem azonosítható a jelentett halálozással), más része azonban matematikai: mi lesz a jó extrapoláció a múltbeli adatokból? Ez a kérdés egyáltalán nem irreleváns, hiszen ha magasabb értéket becslünk, akkor kisebb többlethalálozást kapunk és fordítva. Előadásomban körbejárom azokat a statisztikai módszereket, mellyel ezek az előrejelzések elvégezhetőek, bemutatom a szisztematikus vizsgálatuk módszereit, és az ezekkel kapott eredményeket is. Kitérek arra, hogy ezen eredményekből mi szűrhető le, mi a helyes munkamódszer, és végül rámutatok a megállapítások néhány gyakorlati alkalmazására és jelentőségére.

Fejlett RAG: egy termék és szolgáltatás fejlesztésének néhány tapasztalata

LONDON ANDRÁS

Szegedi Tudományegyetem

A RAG (Retrieval-Augmented Generation) a nagy nyelvi modellek (LLM-ek) feladat-specifikus alkalmazhatóságát hivatott javítani „külső” tudásbázisok segítségével, lehetővé téve a kontextusfüggő és pontosabb válaszadást. A fejlett RAG módszerek területén végzett kísérleteink néhány új megközelítést vizsgálnak a paradigma hatékonyságának javítására. A módszereink között szerepel többek közt tudásgráfok, speciális LLM ágensek és hierarchikus klaszterezési eljárások alkalmazása. Néhány esettanulmányon keresztül bemutatjuk, hogy ezen módszerek hogyan növelik a teljesítményt különböző alkalmazásokban, különösen kontextus kinyerés, illetve a kérdés-megválaszolás tekintetében.

Közös munka: Dombi Gergely, Székely-Kádár Márta, Vincze Sándor (Gear Technologies – kodeage.ai)