

Mélytanuló megoldások üzleti felhasználása



Hol lehet alkalmazni mélytanuló megoldásokat?



- Ahol képesek vagyunk egy jelenséget numerikusan ábrázolni
- Ahol van elegendő jó minőségű tanítóadat
- Ahol garantálni tudjuk az adatok sértetlenségét



Hol érdemes alkalmazni mélytanuló megoldásokat?



- Ahol az egyszerűbb megoldások nem működnek megfelelő hatékonysággal (szabály alapú megoldások, nem neurális hálózat alapú megoldások)

Hol érdemes alkalmazni mélytanuló megoldásokat szerintünk?

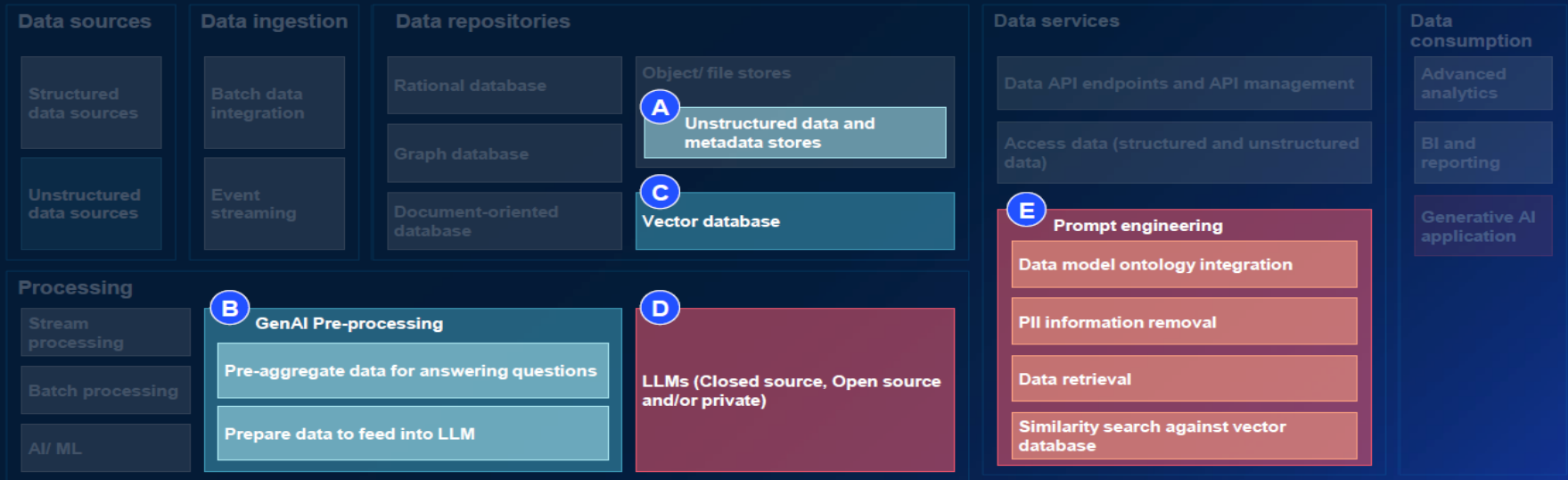
- A nagy nyelvi modellek alkalmazásával újra kell értelmezni, mit értünk hatékony megoldás alatt
- Az ügynök alapú megközelítéssel eszközöket adunk a modellnek, mely segítségével komplex problémákat oldhat meg
- Támogató megoldásokat érdemes létrehozni, ahol a neurális háló segíti az embert, de a felelősség az embernél marad (ügyfélkiszolgálás, csalás megelőzés, ügyfél irritáció észlelése, kérdezz/felelek megoldások)

A komplexitás folyamatosan növekszik



2. Build scale into the data architecture

■ GenAI extensions, with mature tooling/solutions ■ GenAI extensions, with novel/emerging tooling/solutions



E. Prompt engineering

Provide effective prompt engineering fed with context to generate best responses and performance improvement

Hogyan építhetünk megbízható megoldásokat?



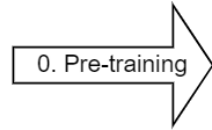
- Data governance (ismerd az adataid)
- Dataset versioning (tudd, hogy mikor milyen állapotát használod)
- Model versioning (tudd, hogy a modell milyen típusú és verziójú adaton tanult)
- Feature store
- Model store

Nagy nyelvi modellek tanítóadata

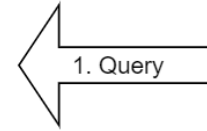
- ELTE – DH-Lab együttműködés keretében épülő magyar nyelvi corpus (~40 milliárd szó)
- Hálózati szinten fizikailag szeparált környezet
- Tanítás indításakor olyan helyre kerülnek az adatok, mely csak és kizárólag a modell tanítását végző folyamat tudja olvasni

Nagy nyelvi modellek használata

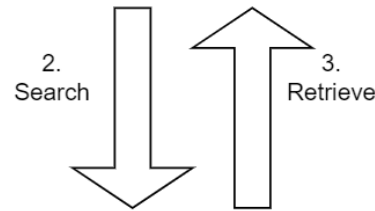
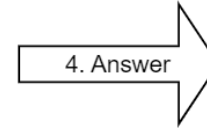
- Külső és belső ügyfél kiszolgálás (tudásbázis, Q&A botok)
- Kisebb, 1.5 milliárd paraméteres modellek kiválóak szöveg klasszifikációs feladatokra (CPU alapú infrastruktúrán is jól tud futni)
- Nagyobb, 13 milliárd paraméteres modellek néhány mintát kapva is képesek nagyon jól teljesíteni, komplexebb feladatokra is alkalmazhatók (ügynök alapú architektúra)
- Retrieval Augmented Generation megoldások (nem a modell tárgyi tudásából dolgozunk ekkor)



Pre-trained LLM



User



Köszönöm a figyelmet!