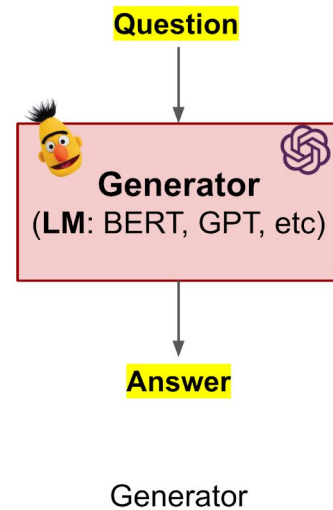
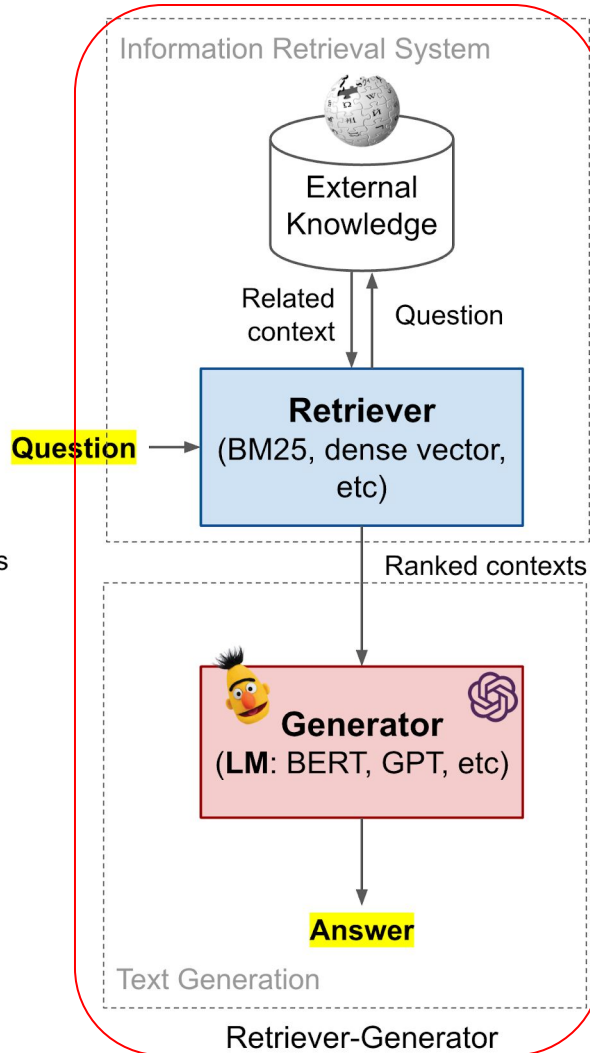
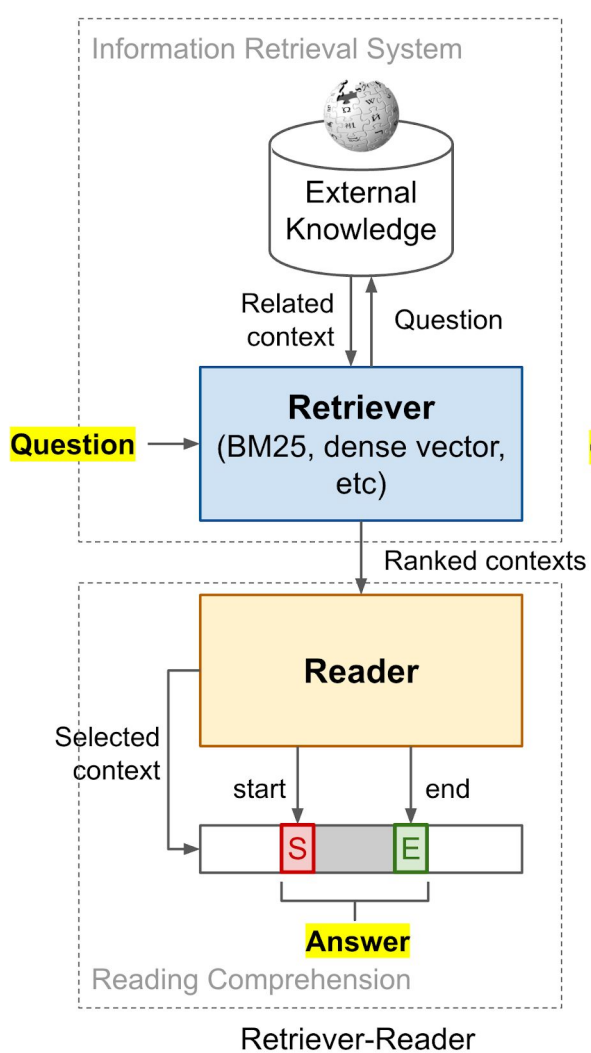
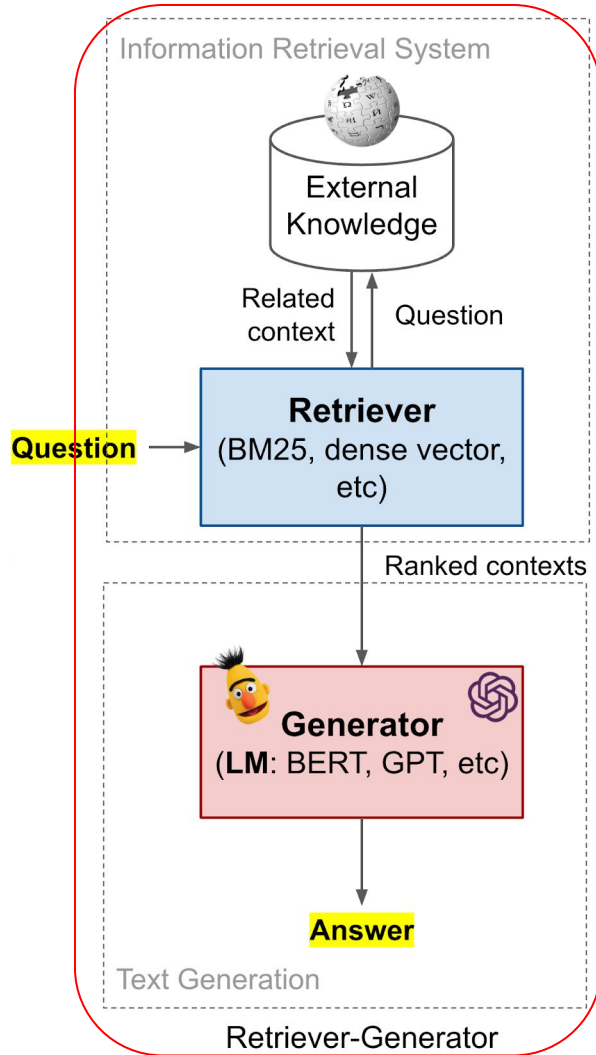


RAGozni a magyart





Retrieval

- Releváns dokumentumok (dokumentum részletek) kiválasztása
- Gyors
- Sok szövegből indul ki

MILQA

- 23700 kérdés (17000 megválaszolható, 6700 nem megválaszolható)
- Wikipedia
- ~Bekezdések + kézzel írt kérdések
- Válasz típusok:
 - Szöveges
 - Rövid (extraktív)
 - Hosszú (extraktív)
 - Kézzel átalakított válasz (absztraktív)
 - Igen/nem
 - Listás
 - Aritmetikai

MILQA - Példa

Kontextus: A második világháború után lassan újra fejlődésnek induló dunai hajózás számára továbbra is szűk keresztmetszetet jelentett a Vaskapu-szoros korlátozott átbocsátó képessége. Ugyanakkor a szurdokok nagyon kedvező lehetőségeket kínálnak vízenergia nyerésére. Mivel a folyó e szakaszon két állam határát alkotja, szükség volt Románia és az akkori Jugoszláv Szocialista Szövetségi Köztársaság együttműködésére a tervek valóra váltására. A Vaskapu I erőmű 1964 és 1972 között épült. Elkészültek a világ legnagyobb vízerőművei közé tartozott 2×2052 MW kapacitásával. Az építményekbe csaknem 3 millió köbméter betont építettek be. A fő gát 441 méter hosszú és 60 méter széles. A rendszer 33 méteres magasra duzzasztott 2 milliárd köbméter vizet. A megemelkedett vízszint miatt a felsőbb Duna-szakasz képe jelentősen megváltozott, a hajózás természetes akadályai viszont megszűntek.

Kérdés: Mennyi betont építettek be a Vaskapu I erőműbe?

Rövid válasz: 3 millió köbméter

Módosított rövid válasz: 3 millió köbmétert

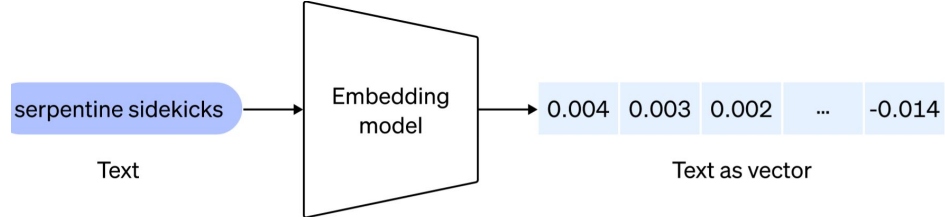
Hosszú válasz: Az építményekbe csaknem 3 millió köbméter betont építettek be.

MKQA

- 26 nyelv
- 10000 kérdés-válasz pár
- Rövid válaszok
 - Sok esetben WikiData entitások
 - Egyébként kézzel fordítva
- Kézzel fordított kérdések
- Keresznyelvi kiértékelés

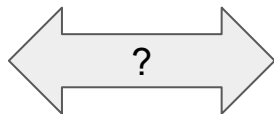
Language	Language Code	Question	Answer and Aliases	Answer Type	Entity ID
English		who produces the most pistachios in the world	Iran, Republic of Iran, ir, Islamic Republic of Iran, Persia, Islamic Rep. Iran	Entity	Q794
Arabic	ar-SA	من أكبر منتج للفستق في العالم	إيران، إيران، الجمهورية الإسلامية الإيرانية، إيران، الجمهورية الإسلامية في إيران، الجمهور الإسلامي الإيرانية، جمهورية إيران الإسلامية، جمهورية إيران الإسلامية، جمهورية إيران، الجمهورية الإسلامية الإيرانية، جمهورية إيران الإسلامية، جمهورية إيران الإسلامية، إيران الأكبر		
Danish	da-DK	hvem producere flest pistacienødder i verden	Iran		
German	de-{DE,AT,CH}	Wer produziert weltweit die meisten Pistazien	Iran, Islamische Republik Iran		
Spanish	es-ES	¿Quién produce la mayor cantidad de pistachos del mundo?	Irán, República Islámica de Irán, Iran, Republica Islamica de Iran		
Finnish	fi-FI	Kuka tuottaa eniten pistaasipähkinöitä maailmassa	Iran		
French	fr-FR	Qui produit le plus de pistaches dans le monde	Iran, Perse, République islamique d'Iran		
Hebrew	he-IL	מי מייצר הכי הרבה פיסטוקים בעולם	איראן		
Hungarian	hu-HU	Ki termeli a legtöbb pisztáciát a világon?	Irán, Iráni Iszlám Köztársaság, Perzsia		
Italian	it-IT	Chi produce più pistacchi al mondo	Iran, Repubblica Islamica dell'Iran		
Japanese	ja-JP	世界で一番ピスタチオを生産しているのは誰ですか	イラン, イラン イスラム共和国		
Khmer	km-KH	អ្នកណាផលិត pistachios ច្រើនជាងគេបំផុតនៅលើពិភពលោក?	អ៊ីរ៉ង់		
Korean	ko-KR	전 세계에서 누가 가장 많은 피스타치오를 생산 하나요	이란		
Malay	ms-MY	siapa menghasilkan pistachios paling banyak di dunia	Iran		
Dutch	nl-NL	wie produceert de meeste pistachio nootjes ter wereld	Iran, Islamitische Republiek Iran, IR, IRN		
Norwegian	nb-NO	hvem lager mest pistasj nøtter i verden	Iran		
Polish	pl-PL	kto produkuje najwięcej pistacji na świecie	Iran		
Portuguese	pt-BR	quem produz mais pistaches no mundo	Irão, ایران، República Islâmica do Irão, Republica Islamica do Irao		
Russian	ru-RU	кто в мире производит больше всего фисташек	Иран, Исламская Республика Иран, Персия, ИРИ, Iran, Иран (Исламская Республика), Название Ирана		
Swedish	sv-SE	Vem producerar mest pistagenötter i världen?	Iran		
Thai	th-TH	ใครผลิตถั่วพิสตาชิโอมากที่สุดในโลก	ประเทศอิหร่าน		
Turkish	tr-TR	Dünyanın en çok antep fıstığını kim üretiyor?	İran, İran İslam Cumhuriyeti		
Vietnamese	vi-VN	nước nào sản xuất nhiều quả hồ trăn nhất thế giới	Iran		
Chinese (Simplified)	zh-CN	谁生产了世界上最多的开心果	伊朗		
Chinese (Hong Kong)	zh-HK	誰是世界上開心果產量最高的人	伊朗, 伊朗伊斯蘭共和國		
Chinese (Traditional)	zh-TW	全世界開心果的主要生產者是誰	伊朗		

Retriever - Embedding



- Kérdés - szövegrészlet párok hasonlósága

Mennyi betont építettek be a Vaskapu I erőműbe?



A második világháború után lassan újra fejlődésnek induló dunai hajózás számára továbbra is szűk keresztmetszetet jelentett a Vaskapu-szoros korlátozott átbocsátó képessége. Ugyanakkor a szurdokok nagyon kedvező lehetőségeket kínálnak vízenergia nyerésére. Mivel a folyó e szakaszon két állam határát alkotja, szükség volt Románia és az akkori Jugoszláv Szocialista Szövetségi Köztársaság együttműködésére a tervek valóra váltására. A Vaskapu I erőmű 1964 és 1972 között épült. Elkészültek a világ legnagyobb vízerőművei közé tartozott 2×2052 MW kapacitásával. Az építményekbe csaknem 3 millió köbméter betont építettek be. A fő gát 441 méter hosszú és 60 méter széles. A rendszer 33 méteres magasra duzzasztott 2 milliárd köbméter vizet. A megemelkedett vízszint miatt a felsőbb Duna-szakasz képe jelentősen megváltozott, a hajózás természetes akadályai viszont megszűntek.

Embedding

- Végtelen modell HunggingFace-en
- Problémák magyar retrieval fronton:
 - Rengeteg rendszer nem tud magyarul
 - Olyan többnyelvű modell, amit kérdés - válasz párokon is finomhangoltak
 - Nem elég, ha csak kér szöveg hasonlóságát írja le
 - Nincs elég magyar adat finomhangolásra, ezért jellemzően angol adaton finomhangolt többnyelvű modellek kellene

Embedding

- Végtelen modell HunggingFace-en
- Problémák magyar retrieval fronton:
 - Rengeteg rendszer nem tud magyarul
 - Olyan többnyelvű modell, amit kérdés - válasz párokon is finomhangoltak
 - Nem elég, ha csak kér szöveg hasonlóságát írja le
 - Nincs elég magyar adat finomhangolásra, ezért jellemzően angol adaton finomhangolt többnyelvű modellek kellene

mContriever	2021	512	mBERT	80,87
mE5	2023	512	XLM-Roberta	87,84
BGE-M3	2024	8192	XLM-Roberta	87,67
OpenAI embedding v3 - large	2024	8192	???	87,79

Szóalak alapú módszerek 2024-ben

- BM25
- A szövegek hossza nem igazán probléma.
- Csak a szóalakokra épít
 - érzékeny a typo-kra: erre megoldás a fuzzy search
 - érzékeny az eltérő megfogalmazásra
- Előfeldolgozás
 - HuSpaCy
 - Lemmatizálás (szótövezés)
 - söröket -> sör
 - Szófaj alapú szűrés
 - pl: kérdőszavak

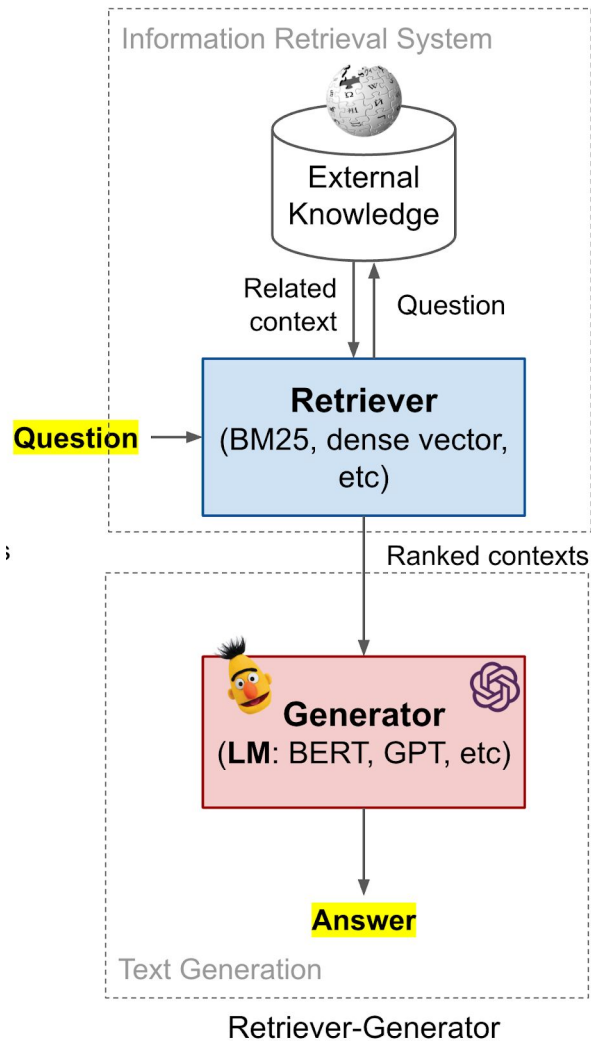
Hybrid módszerek 2024-ben

- BM25 és embedding alapú megoldás keveréke
- A score-ok kombinálása mContriever esetén még javított
- BGE-M3 és mE5 esetén már nem

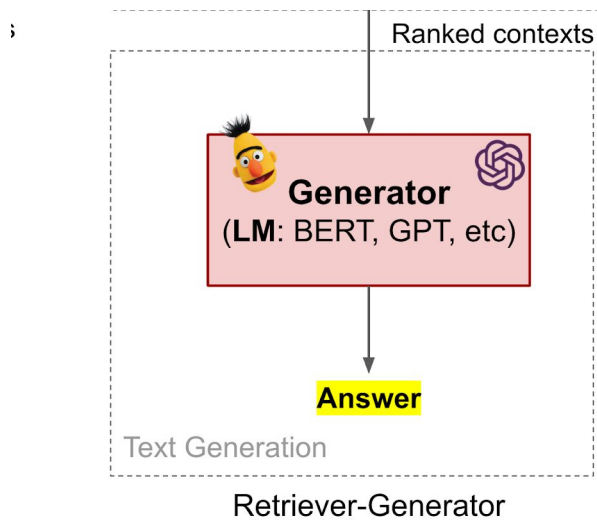
Ranker

- Nyelvi modell hasonlítja össze a legjobb találatokat a kérdéssel
- Pontosabb sorbarendezése
- Lassabb
- Hasonló méretű modellek, mint az embeddingeknél (jellemzően sokkal kisebb, mint amiket szöveg generálásra használunk)

Generálás



Generálás





PHASE 1

PHASE 2

PHASE 3

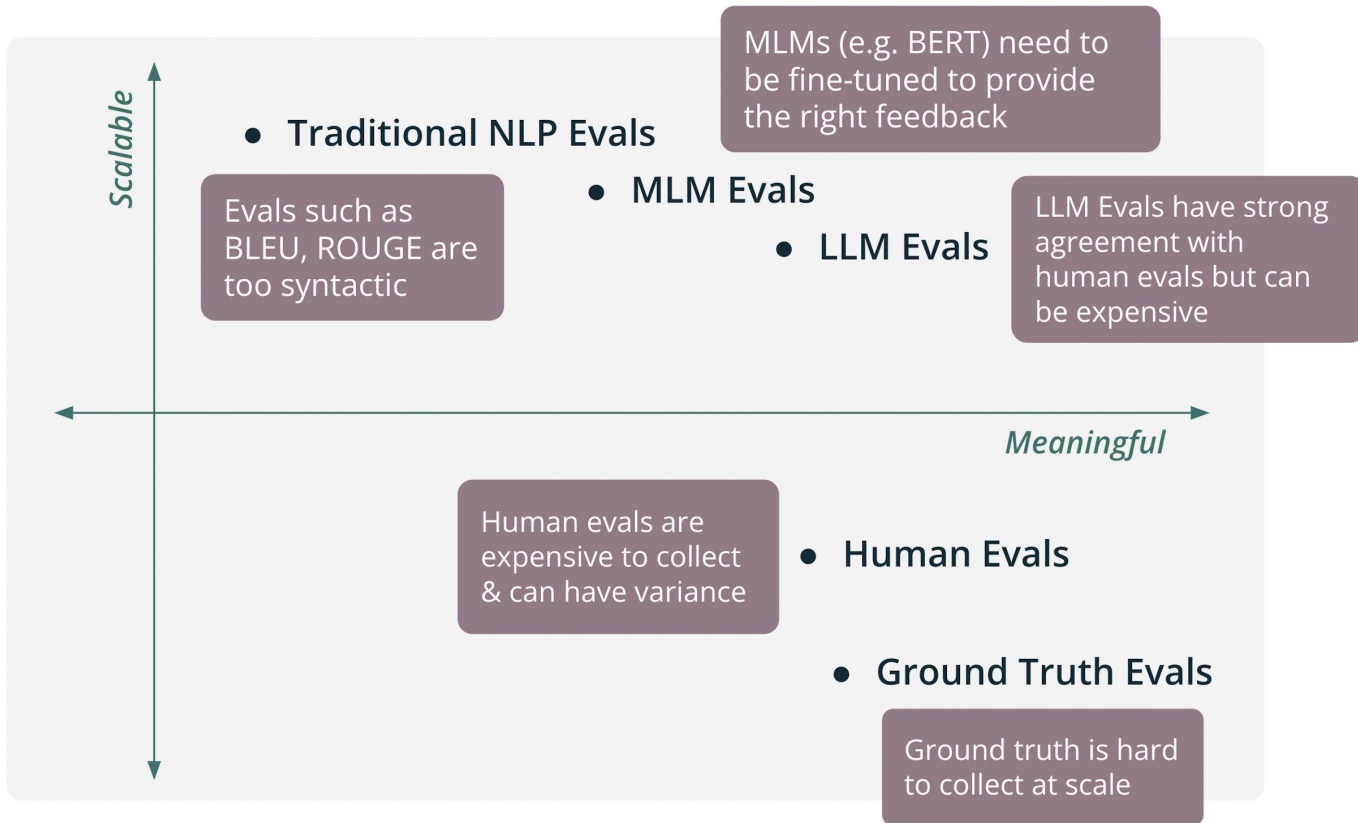
LLM + prompt



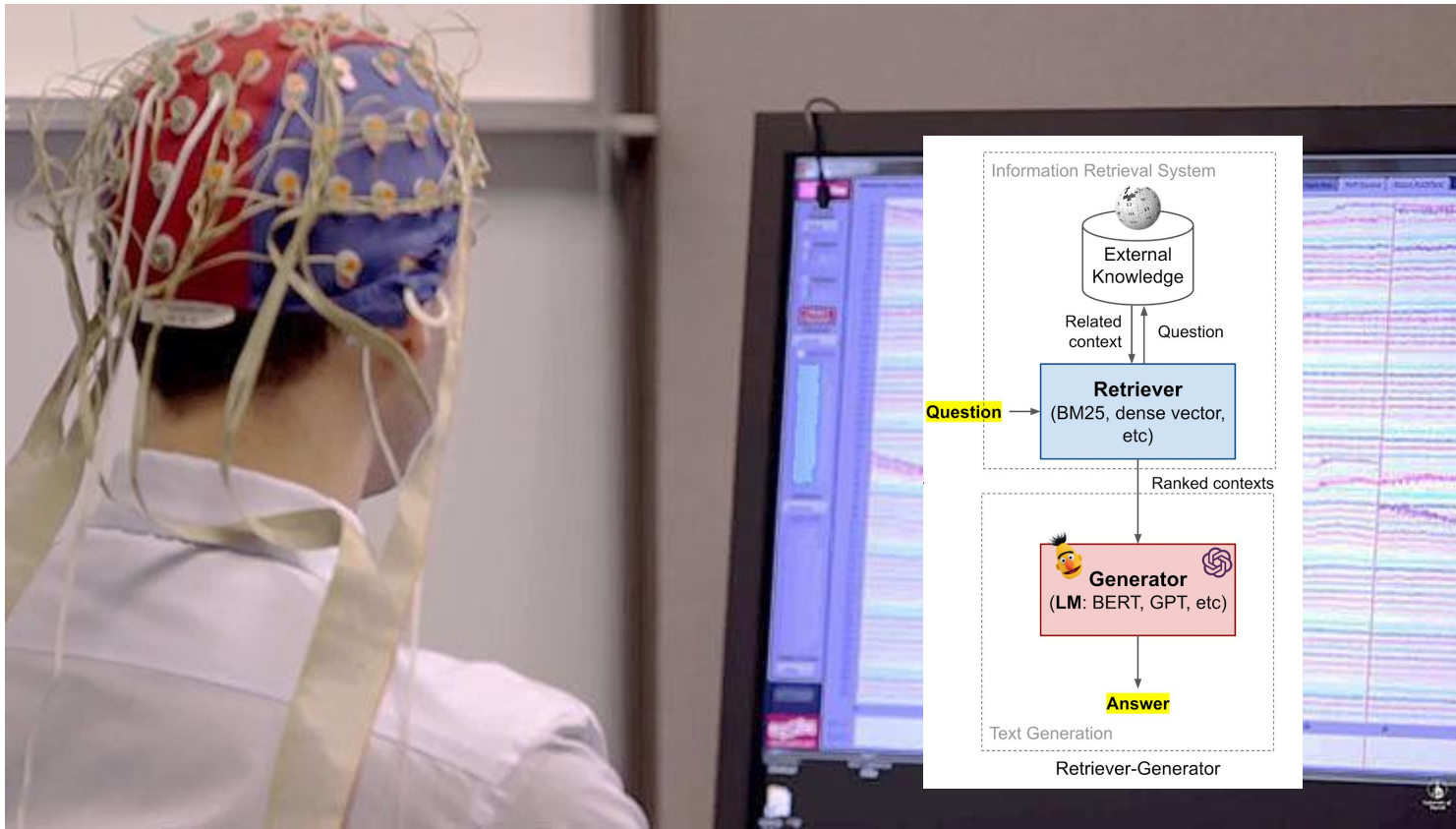
Profit



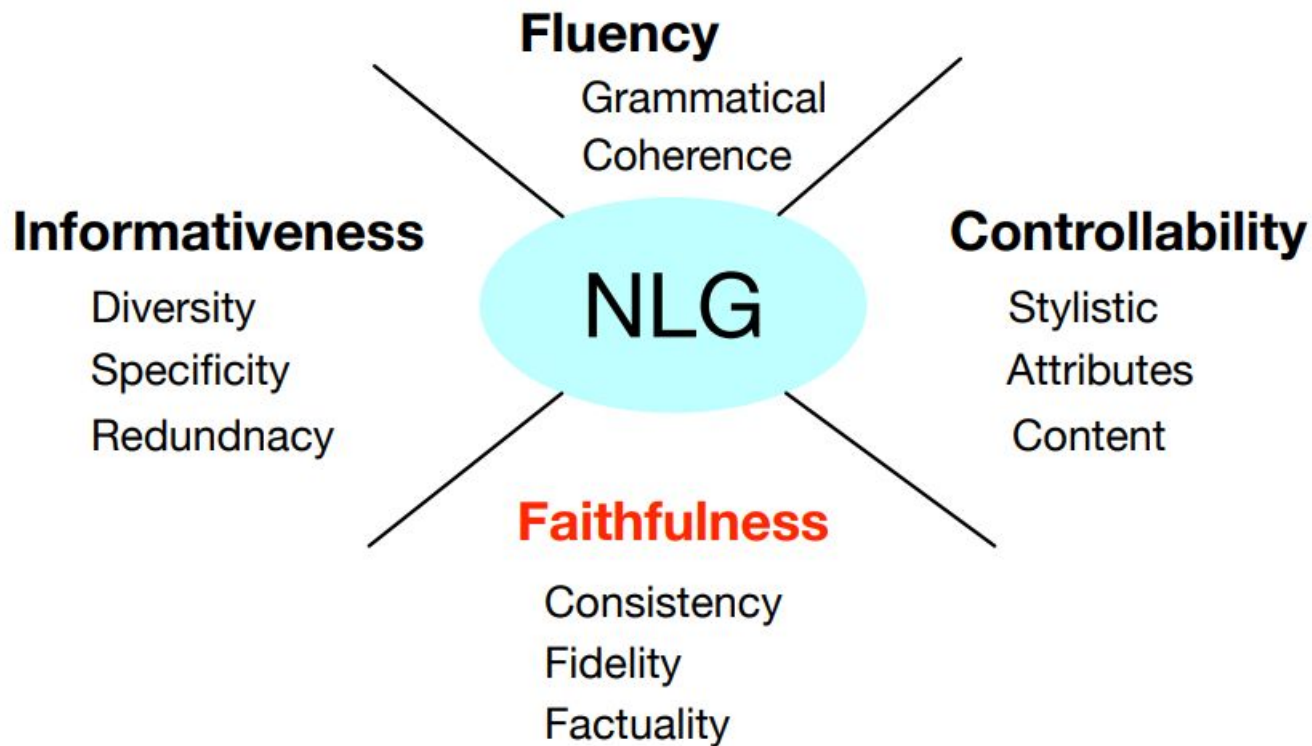
Hogyan értékeljük ki a RAG-ot?



Emberi kiértékelés



Kiértékelés szempontjai



Automatikus kiértékelés

Előnyök:

- Nem igényel emberi beavatkozást
 - Olcsóbb
 - Gyorsabb
- Látványos hibákat azonnal jelezni tud
 - Smoke test
- Segítheti az emberi elemzést
 - Például csak azokkal a példákkal foglalkozunk amikre a rendszer alacsony bizonyossággal válaszolt

Automatikus kiértékelés

Hátrányok:

- Kevésbé pontos, kis különbségek kimutatására nem alkalmas.
- Meta kiértékelés: Metrika tesztelése az általunk választott feladatra!!!

Correctness

A generált válasz mennyire hű a már ismert válaszhoz.

Kérdés: Melyik a harmadik legnagyobb magyar város?

Jó válasz: Szeged

Generált válasz: A harmadik legnagyobb magyar város Szeged. Az ország délkeleti részén, a Tisza folyó partján található, és Csongrád-Csanád megye székhelye.

Faithfulness

A generált válasz mennyire hű a kapott kontextusokhoz.

Kérdés: Melyik a harmadik legnagyobb magyar város?

Jó válasz: ???

Kontextus:

Városok listája:

Budapest - 1 685 342

Debrecen - 199 858

Szeged - 158 797

A világ harmadik legnagyobb városa Shanghai.

A magyar válogatott harmadik lett a csoportkörben.

Generált válasz: A harmadik legnagyobb magyar város Szeged. Az ország délkeleti részén, a Tisza folyó partján található, és Csongrád-Csanád megye székhelye.

Faithfulness

A generált válasz mennyire hű a kapott kontextusokhoz.

- A teszteléshez csak kérdésekre van szükség.
- A retriever hibáját nem teszteli.

Faithfulness - hiba típusok

The first vaccine for Ebola was approved by the FDA in 2019 in the US, five years after the initial outbreak in 2014. To produce the vaccine, scientists had to sequence the DNA of Ebola, then identify possible vaccines, and finally show successful clinical trials. Scientists say a vaccine for COVID-19 is unlikely to be ready this year, although clinical trials have already started.

Categorization		Examples	
Intrinsic Error	Semantic Frame Errors	Predicate Error (PredE)	The Ebola vaccine was rejected by the FDA in 2019.
		Entity Error (EntE)	Scientists say a vaccine for Ebola is unlikely to be ready this year.
		Circumstance Error (CircE)	The first vaccine for Ebola was approved by the FDA in 2014 .
	Discourse Errors	Co-reference Error (CorefE)	The first vaccine for Ebola was approved in 2019. They say a vaccine for COVID-19 is unlikely to be ready this year.
		Discourse Link Error (LinkE)	To produce the vaccine, scientists have to show successful human trials, then sequence the DNA of the virus.
Extrinsic Error	Factual	China has already started clinical trials of the COVID-19 vaccine.	
	Non-Factual	China didn't start clinical trials of the COVID-19 vaccine.	

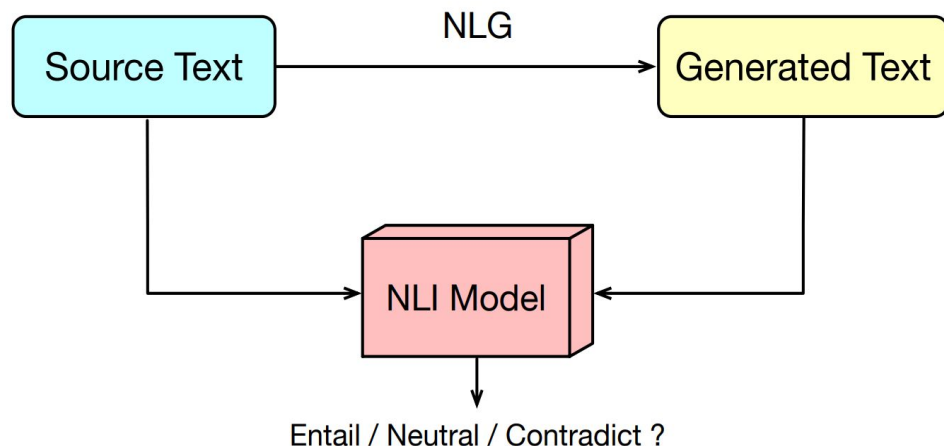
Faithfulness - NLI

Natural Language Inference - Megvizsgálja, hogy egy állítás és egy arra adott hipotézis milyen viszonyban állnak egymással.

- Következmény
- Semleges
- Ellentmondás

Magyar erőforrások:

- Betanított többnyelvű NLI modellek elérhetők a HuggingFace-en.
 - [MoritzLaurer/mDeBERTa-v3-base-xnli-multilingual-nli-2mil7](#)

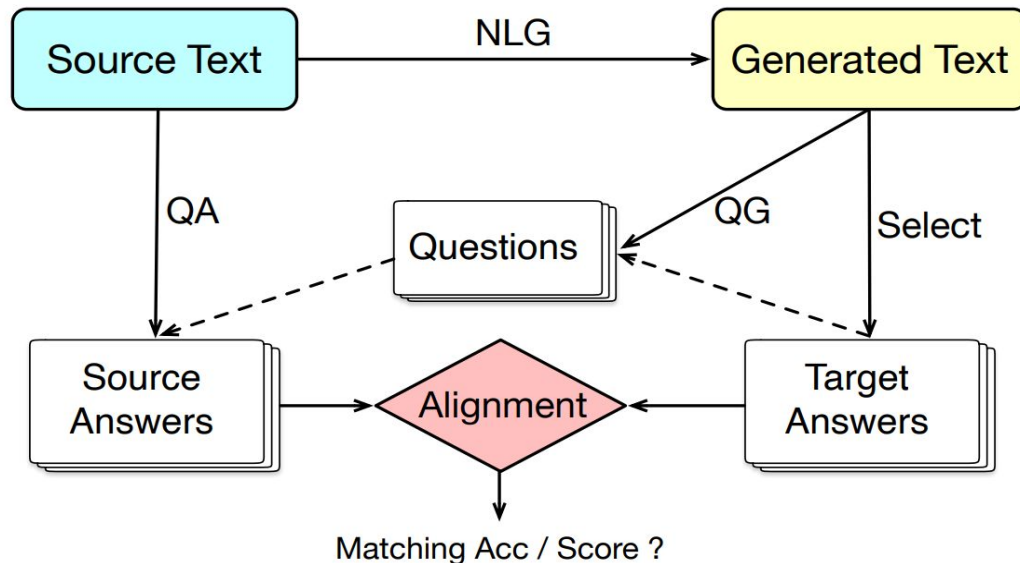


Faithfulness - Question Answering

A generált válasz alapján kérdéseket generálunk. Megnézzük, hogy egy külön rendszer ugyanazt válaszolja-e a forrás alapján, mint a válasz alapján.

Magyar erőforrások:

- Extraktív QA: SHunQA
 - Modellek elérhetők HF-en
- Kérdés generálás
 - Többnyelvű llm-ek



Faithfulness - Factuality

A generált szövegben található Named Entity-k (és kapcsolataik) megtalálhatók a forrás szövegben is?

Magyar nyelvű NER-ek, pl:

- HuSpaCy
- emBERT

A harmadik legnagyobb magyar város **Szeged LOC**. Az ország délkeleti részén, a **Tisza LOC** folyó partján található, és **Csongrád-Csanád megye LOC** székhelye.

Faithfulness - Factuality

A generált szövegben található Named Entity-k (és kapcsolataik) megtalálhatók a forrás szövegben is?

Magyar nyelvű NER-ek, pl:

- HuSpaCy
- emBERT

A harmadik legnagyobb magyar város **London LOC** .

Faithfulness - Prompt

LLM megkérdezése a kontextus és a válasz kapcsolatáról.

- Könnyen használható
- Változatos domaineken is jól működik
- Költségesebb
- Sokszor ugyanaz a modell magát ellenőrzi

Magyarul tudó LLM-ek használhatók a feladatra.



Kérdés generálás

Eldöntendő példák generálása:

Forrás szöveg: *Megerősítéses tanulás:* Szokás külön kategóriának venni a megerősítéses tanulást, ahol az $X \rightarrow Y$ függvény létezik ugyan, de tanuló példák nem adottak direkt módon. A feladat itt az, hogy egy állapottérben az ágens megtanulja az optimális stratégiát (X az állapotok, Y pedig a cselekvések) úgy, hogy a jövőben érintett állapotokban összegyűjtött jutalmak maximalizálását érje el.

Generált “kérdés”: A megerősítéses tanulás során a tanuló példák direkt módon adottak hasonlóan a felügyelt tanuláshoz.

Köszönöm a figyelmet