



KASYS

筑波大学知識獲得システム研究室
UNIVERSITY OF TSUKUBA
KNOWLEDGE ACQUISITION SYSTEM LAB.

AI時代の情報検索技術

筑波大学/国立情報学研究所

加藤 誠

- 2012年9月：京都大学大学院 情報学研究科 博士後期課程 修了
- 2012年10月～2019年3月：京都大学大学院 情報学研究科 特定助教/特定講師
- 2019年4月～現在：筑波大学 図書館情報メディア系 准教授
- 2025年4月～現在：国立情報学研究所 情報社会相関研究系 准教授

- 研究テーマ：情報検索

- 足して2で割る検索
- アナロジーによる検索
- クエリ推薦
- モバイル検索
- 情報検索のオンライン評価
- 統計データ検索
- 大規模言語モデルに基づく検索
- 低資源下における情報検索システム構築



 ビジネス+IT

SearchGPTとは何かを解説、なぜGoogle検索が「もう不要」と言えるのか

しかし、SearchGPTがPerplexityやグーグル検索と同等、またはそれ以上の検索体験を可能にするとすると、様相は大きく変わることになる。ChatGPTユーザーは...

2024/09/25

 クーリエ・ジャポン

AIによって「検索」は不要に グーグルに迫る3つの脅威を米経済紙が指摘

それは、グーグルがこれまで形作り、依存してきたウェブサイトの全体的なエコシステムが劣えていくことだ。AI生成コンテンツの増加により、ウェブの検索...

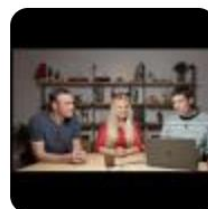
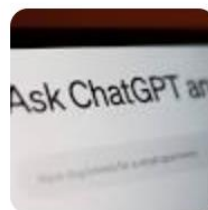
2025/01/03

 ITmedia

AI検索「SearchGPT」リニューアル Google検索いよいよ不要に.....？：その悩み、生成AIが解決（1/2 ページ）

AI検索「SearchGPT」リニューアル Google検索いよいよ不要に.....？：その悩み、生成AIが解決（1/2 ページ）・連載：その悩み、生成AIが解決・著者プロフィール：...

2025/01/03



perplexity

AI時代の情報検索技術は？



<https://www.perplexity.ai/>

Deep Research

What can I help with?

Compile a research report on how the retail industry has changed in the past 3 years. Use bullets and tables where necessary for clarity!

Attach Search Deep research

Create image

Analyze data

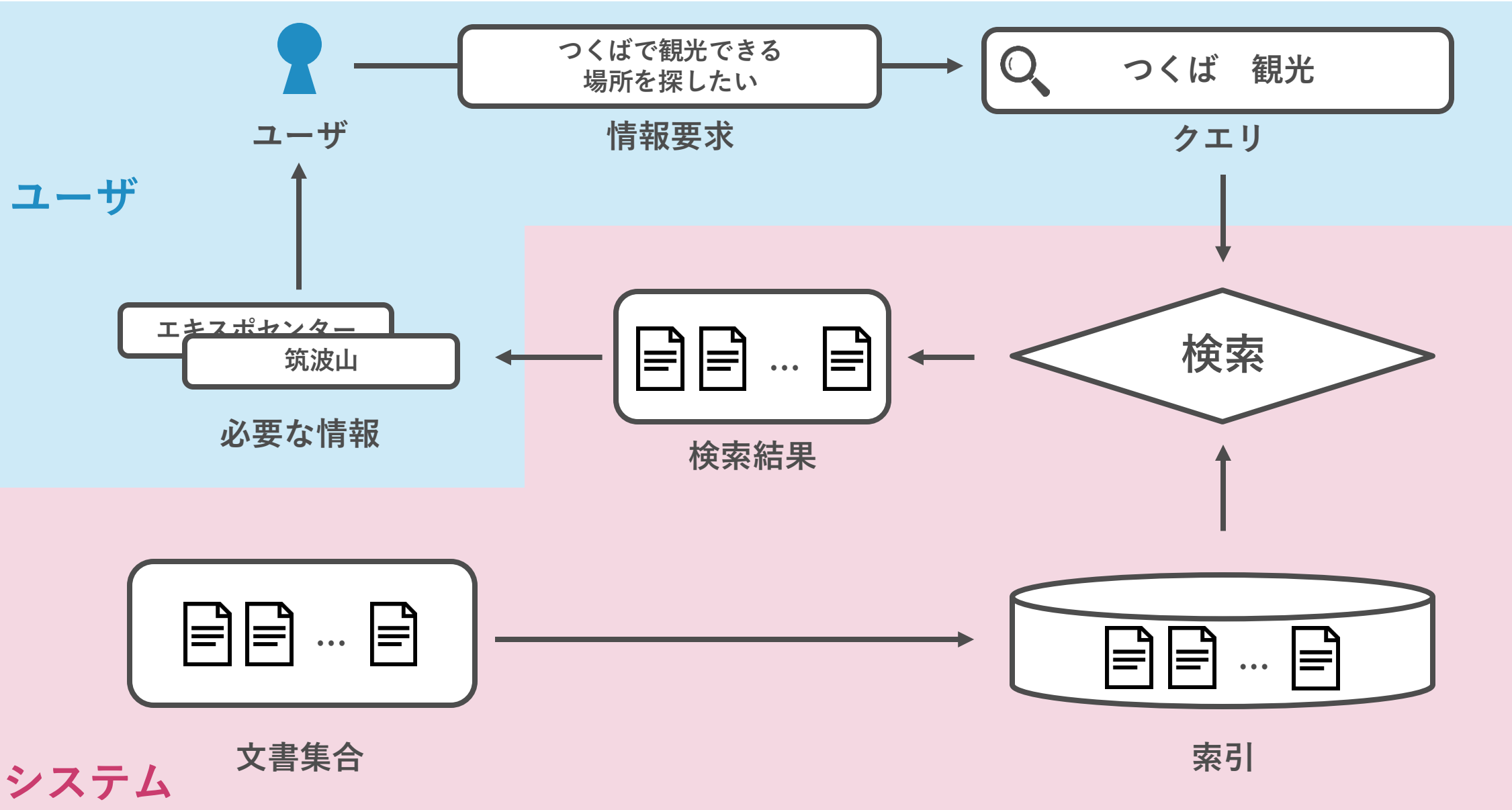
Code

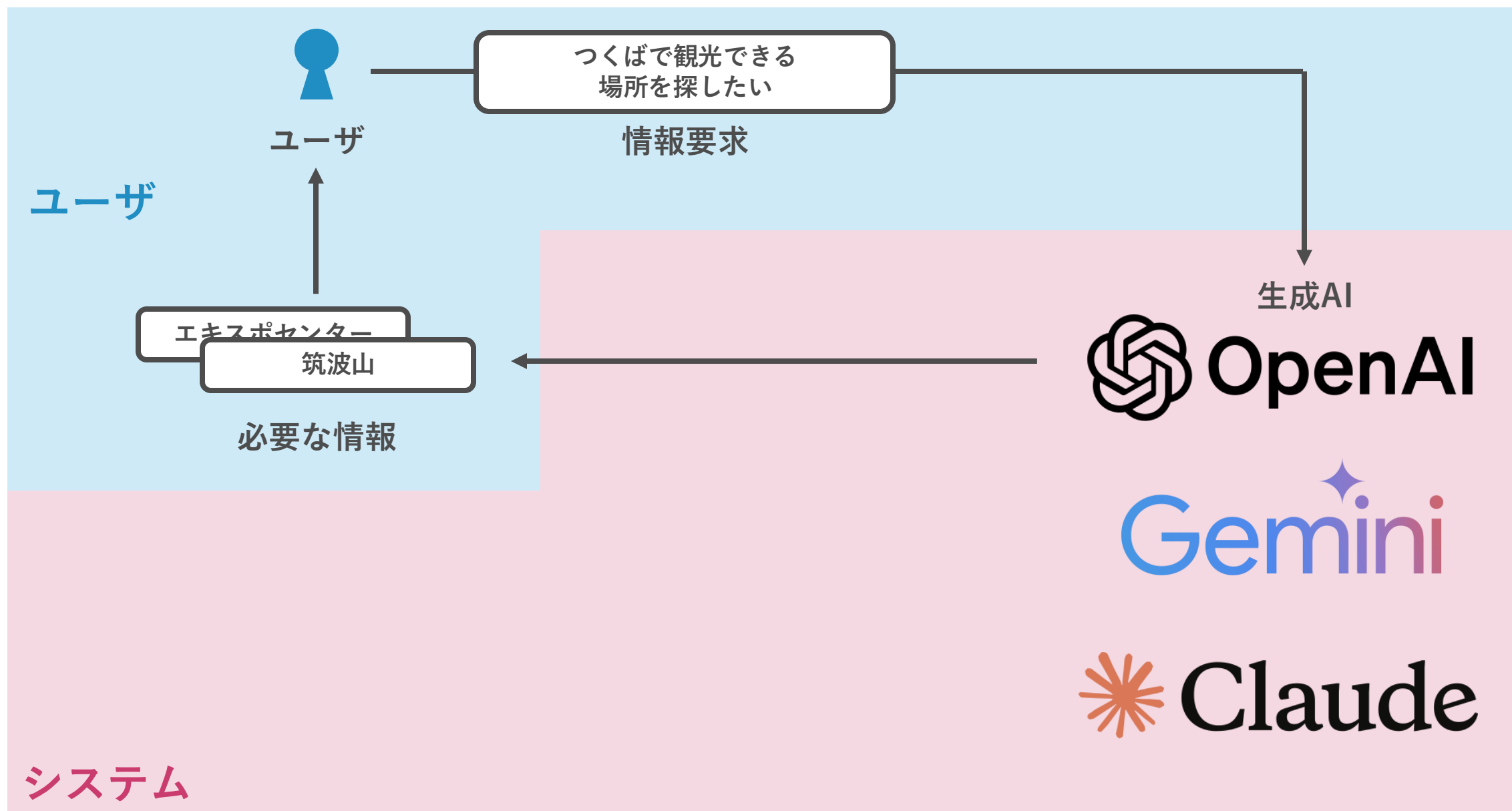
Surprise me

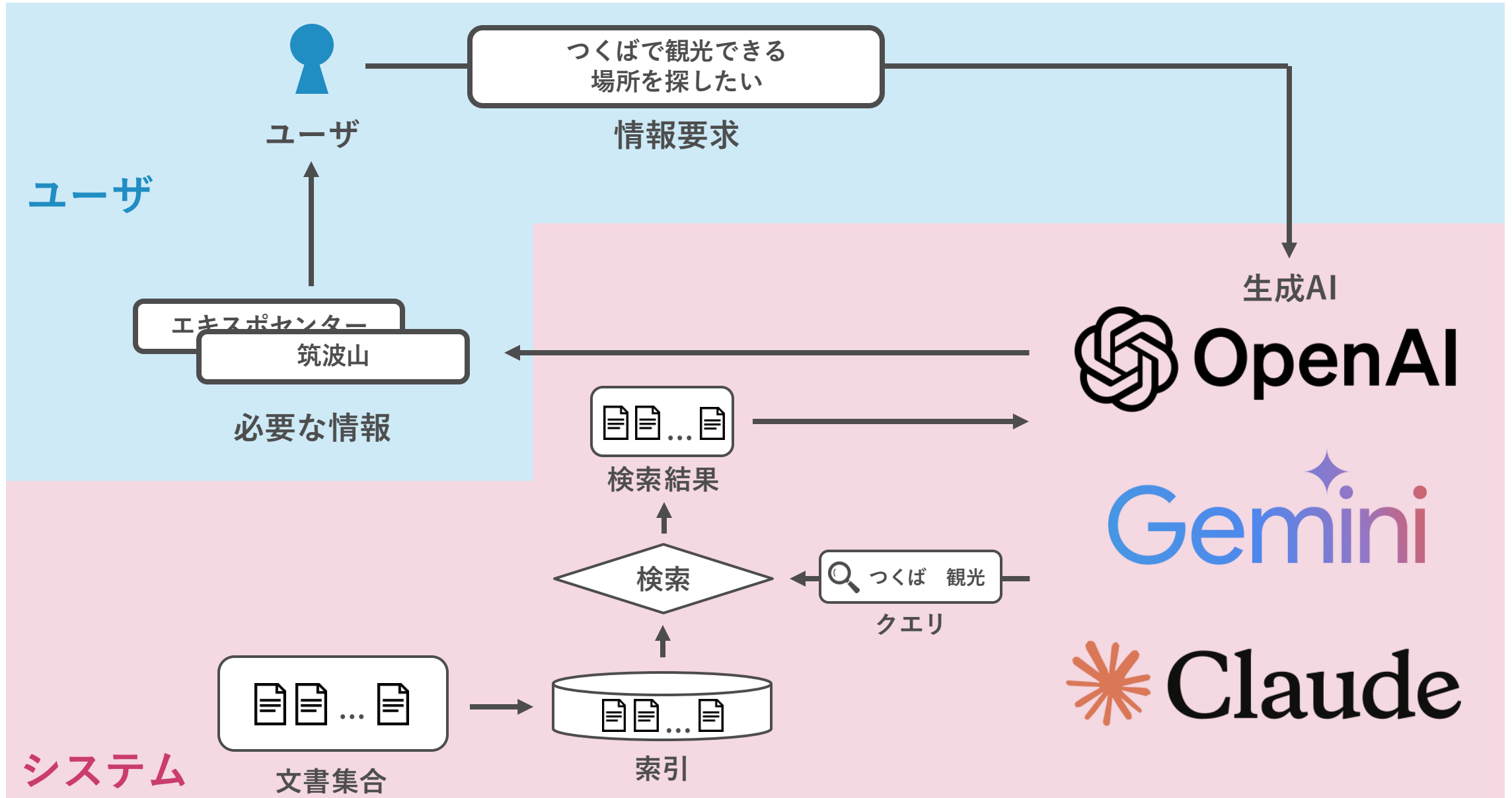
Get advice

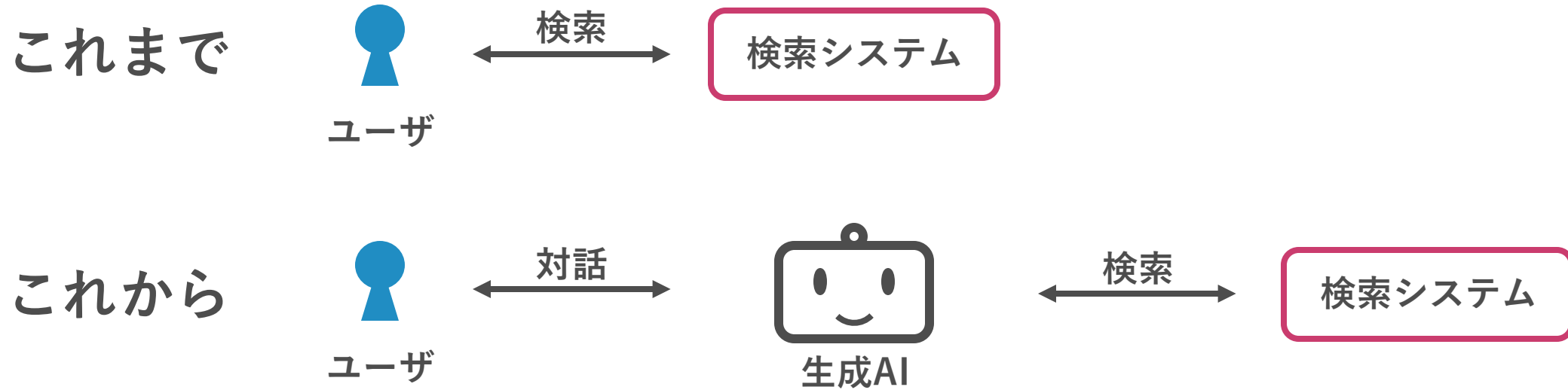
More

<https://openai.com/ja-JP/index/introducing-deep-research/>









- この仕組みを Retrieval-augmented Generation (RAG) と呼ぶ
 - 生成AIが持つ知識では回答を生成できない場合（新規情報やドメイン固有の情報，社外秘の情報など）に外部知識を参照
 - ハルシネーションを抑止する効果もある

生成AIが検索するのであれば、

- ・ Q. 検索システムの精度が悪くても問題ないのでは？
- ・ Q. 人間とは違ってうまく検索してくれるのでは？
- ・ Q. 人間相手と同様に上位に良い結果を提示すればよい？

→ A. いいえ

回答に貢献しない検索結果が生成AIの回答を捻じ曲げる

Wu et al. How Easily do Irrelevant Inputs Skew the Responses of Large Language Models? COLM 2024.

POPQA				
Models	1: 0	3: 0	1: 1	3: 1
GPT-3.5 Turbo	22.5	27.4	1.7	5.5
Llama2-7B	83.5	91.1	11.8	42.4

(回答に貢献しない結果の数)
:
(貢献する結果の数)

表中の値は実際の回答と異なる回答をした割合

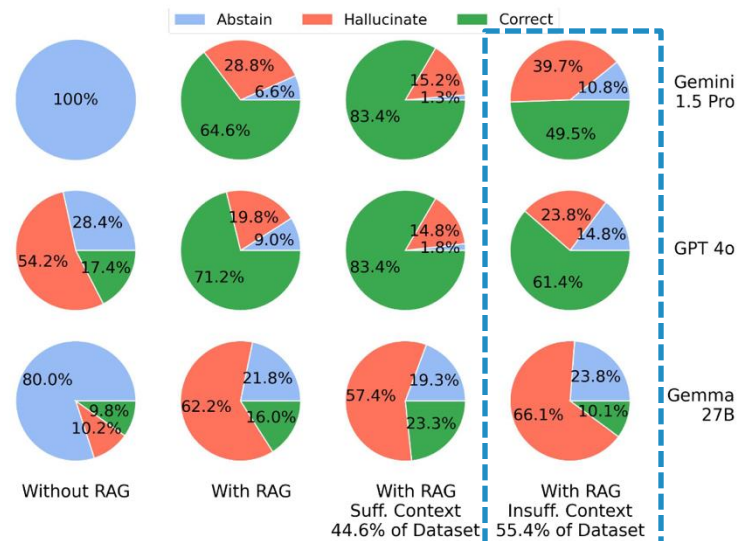
回答を含む検索結果を得られた場合でも、回答に貢献しない結果が多く含まれると、異なる回答をしてしまう

Joren et al. Sufficient Context: A New Lens on Retrieval Augmented Generation Systems. ICLR 2025.

Examples: sufficient context

Question: Who is Lya L. married to?	
Sufficient Context	Context A Lya L. married Paul in 2020... They looked happy together at the recent event.
Sufficient Context	Context B Lya L. - Wikipedia Born: October 1, 1980 Spouse: Paul (m. 2020)
Insufficient Context	Context C Lya L. married Tom in 2006... They divorced in 2014... Lya went on dates with Paul in 2018...
Insufficient Context	Context D Lya L. is an astronaut, born in Ohio.... Lya has two children... Lya's parents are lawyers...

Categorizing Model Responses (Musique Dataset)



回答に必要な検索結果が得られない場合に、回答を放棄（図中■）するのではなく、ハルシネーション（図中■）の割合が増加

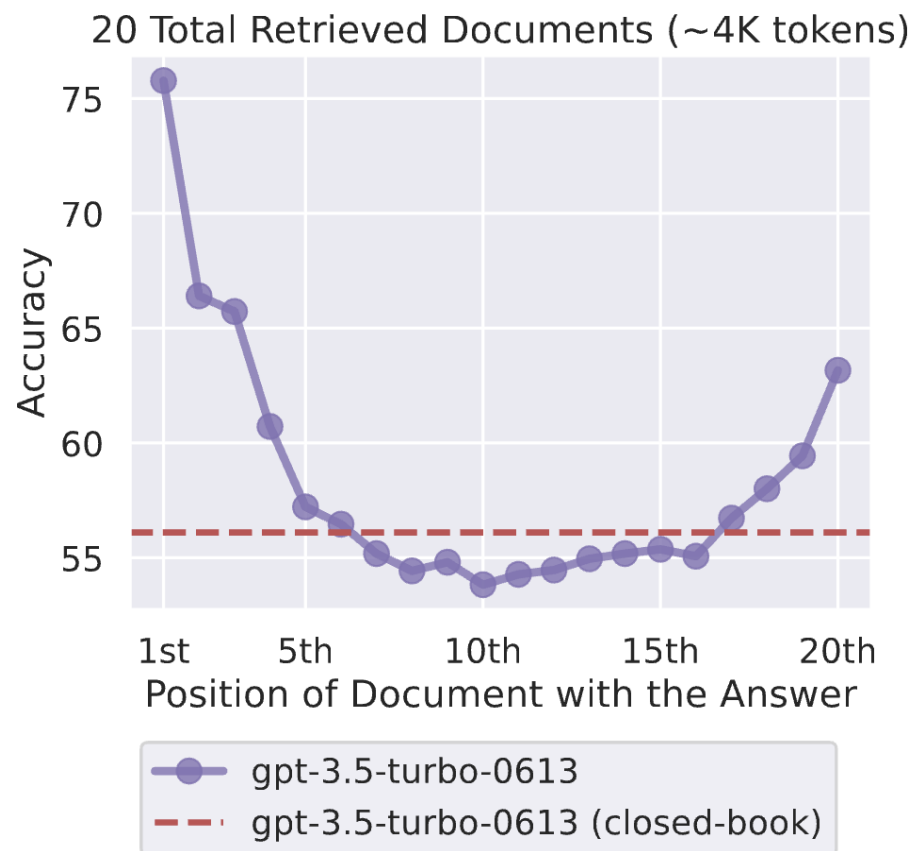
生成AIが生成したクエリ（クエリ拡張）は 生成AI自身がクエリに詳しくないと有効ではない

Abe et al. LLM-based Query Expansion Fails for Unfamiliar and Ambiguous Queries. SIGIR 2025.

Retrieval	Expansion	NQ				TQ				MS MARCO				BioASQ			
		NDCG@10		Recall@100		NDCG@10		Recall@100		NDCG@10		Recall@100		NDCG@10		Recall@100	
		■	✓	■	✓	■	✓	■	✓	■	✓	■	✓	■	✓	■	✓
BM25	No Exp	21.69	19.94	29.63	24.40	22.07	22.45	22.25	21.63	27.24	20.05	68.01	64.38	57.36	39.66	82.54	64.53
	Q2E	+5.98	+7.74	+6.31	+6.54	+1.68	+2.32	+1.81	+2.52	-1.60	+1.59	-0.75	+4.47	+1.51	+0.33	-0.16	+3.19
	Q2D	+7.42	+9.44	+6.67	+7.61	+3.18	+3.93	+2.94	+3.38	-1.13	+4.00	+1.29	+7.78	+1.55	+1.09	+2.16	+2.87
	GaQR (S)	+3.77	+3.80	+3.98	+3.29	+1.63	+1.66	+1.45	+1.85	+0.01	+1.79	+2.17	+4.21	+2.91	+1.47	+1.51	+3.24
Contriever	No Exp	41.72	39.81	47.75	42.31	35.50	39.48	30.89	32.33	44.07	38.62	87.88	89.83	46.43	32.46	69.57	56.14
	Q2E	+1.47	+4.31	+2.03	+3.77	+1.11	+0.90	+1.68	+2.03	-8.63	-4.38	-7.23	-2.88	-7.48	-0.65	-7.10	+2.07
	Q2D	-2.34	+2.52	-0.24	+3.36	-1.92				-14.10	-8.24			-1.89		-9.48	-0.10
	GaQR (S)	+0.55	+2.16	+0.11	+1.18	+0.07	知識なし改善率			-7.13	-3.39	知識あり改善率			-0.40	-7.43	+0.10

検索結果が20件返ってきた場合に，回答に貢献する結果は1-5位，もしくは，17-20位にあると良い

Liu et al. Lost in the Middle: How Language Models Use Long Contexts. TACL 2024.

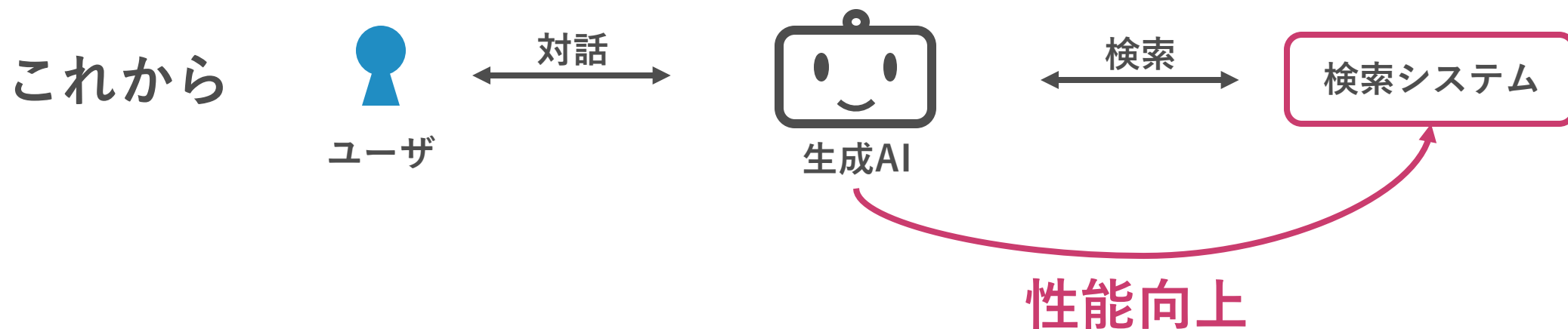


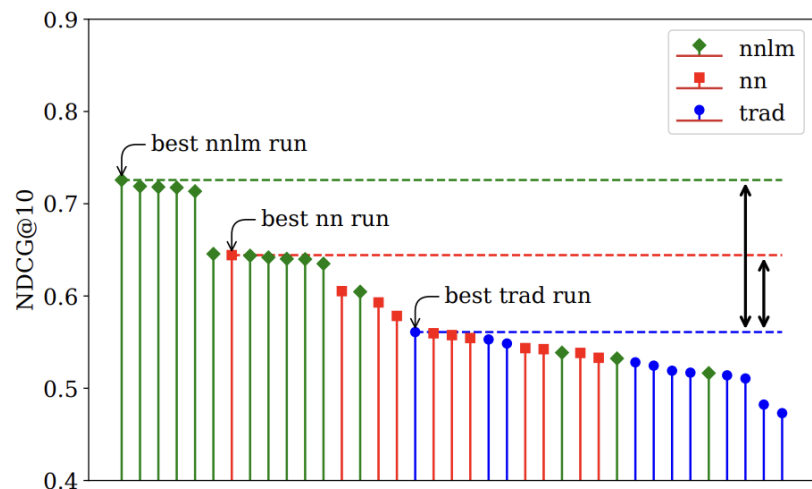
- 回答に貢献する結果の順位を変えたときの回答の正解率
 - 1位，2位，3位，20位，4位，19位，18位，5位，17位の順に正解率が高い
 - 「Lost in the middle」現象と呼ばれる
- ランダムな文書を検索結果に含めたほうが性能が高くなるという報告も

Cuconasu et al. The Power of Noise: Redefining Retrieval for RAG Systems. SIGIR 2024.

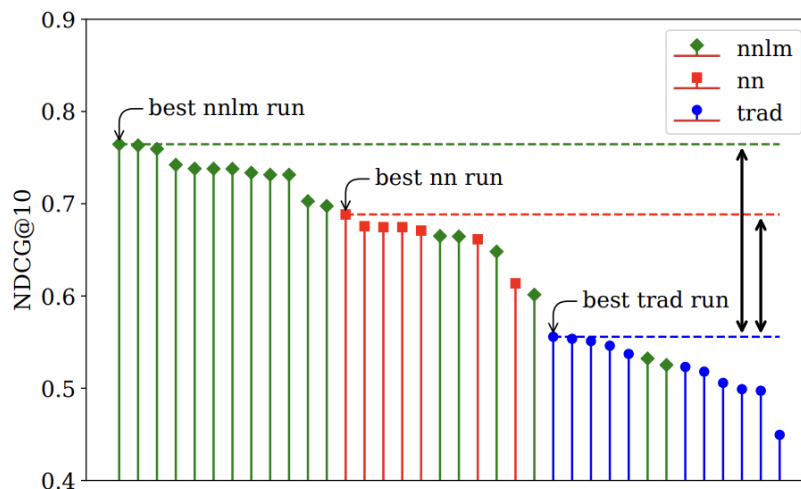
対象が人間から生成AIに変化しつつあるものの
依然として情報検索技術の性能向上が求められている

- (生成) AIによる近年の情報検索技術の発展
- ベクトル検索 (or 密検索) による性能向上
 - LLM-as-a-Judgeによる情報検索技術の評価
 - 合成データによる検索モデルの学習／ドメイン特化





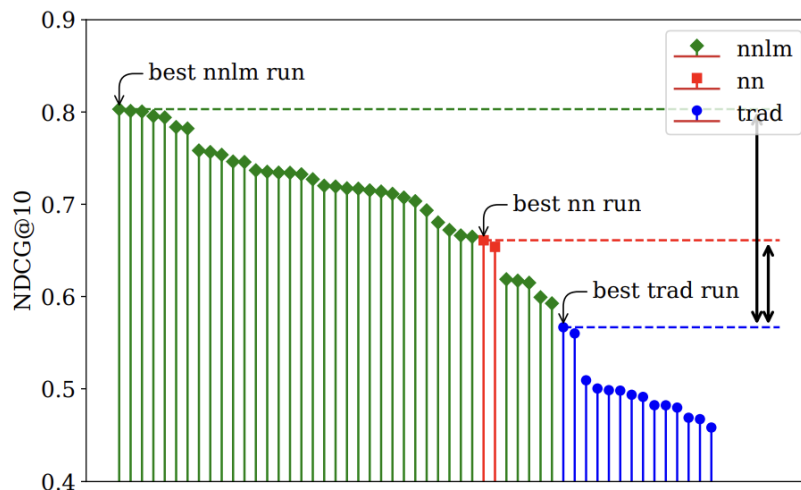
(a) Document ranking task (2019)



(b) Passage ranking task (2019)



(c) Document ranking task (2020)



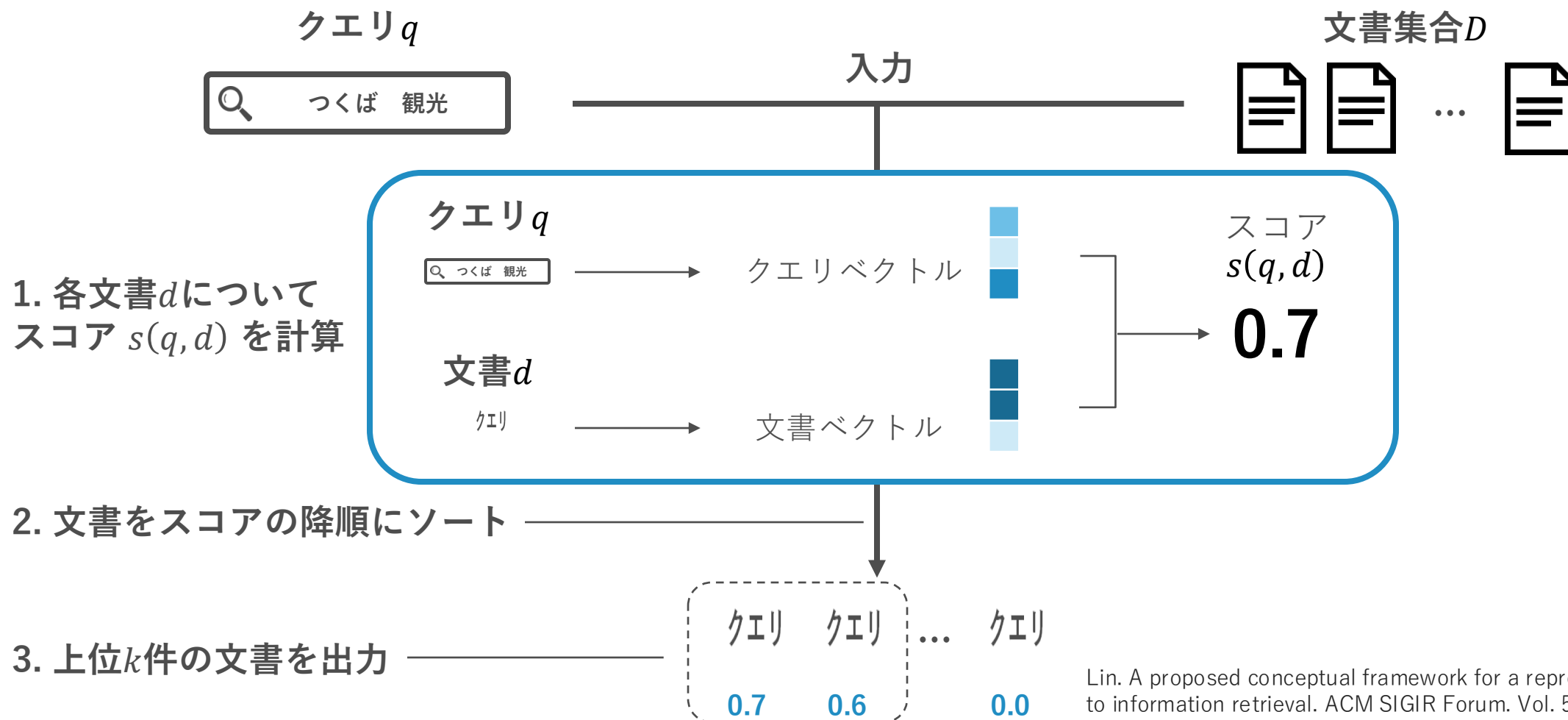
(d) Passage ranking task (2020)

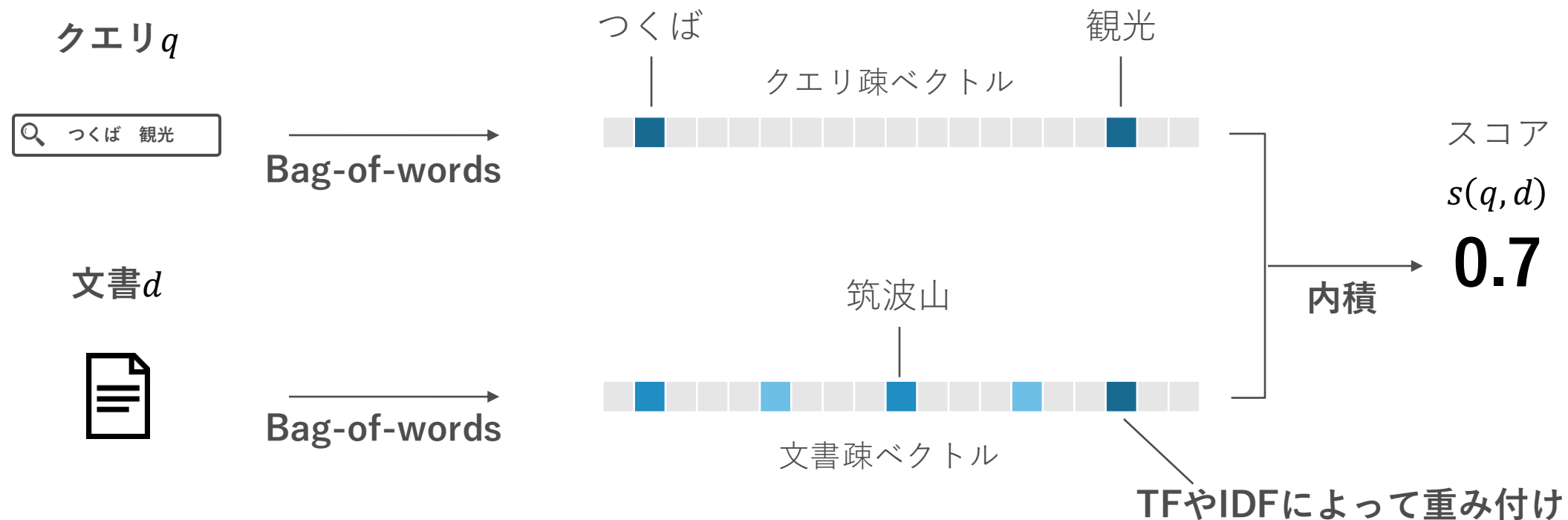
TREC Deep Learningトラック
における大規模言語モデル
(◆nnlm)の性能向上

伝統的モデル (●trad) に
大差をつけている

特にパッセージ検索で顕著

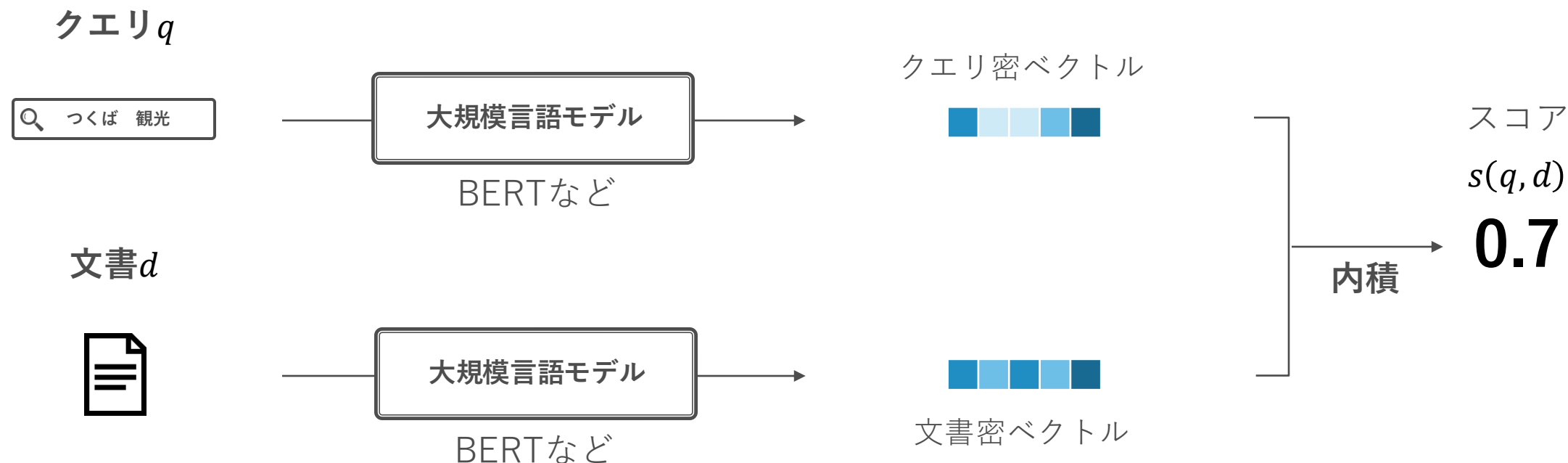
多くの検索アルゴリズムは各文書についてスコア $s(q, d)$ を計算して降順に上位 k 件の文書を選ぶ, と説明可能





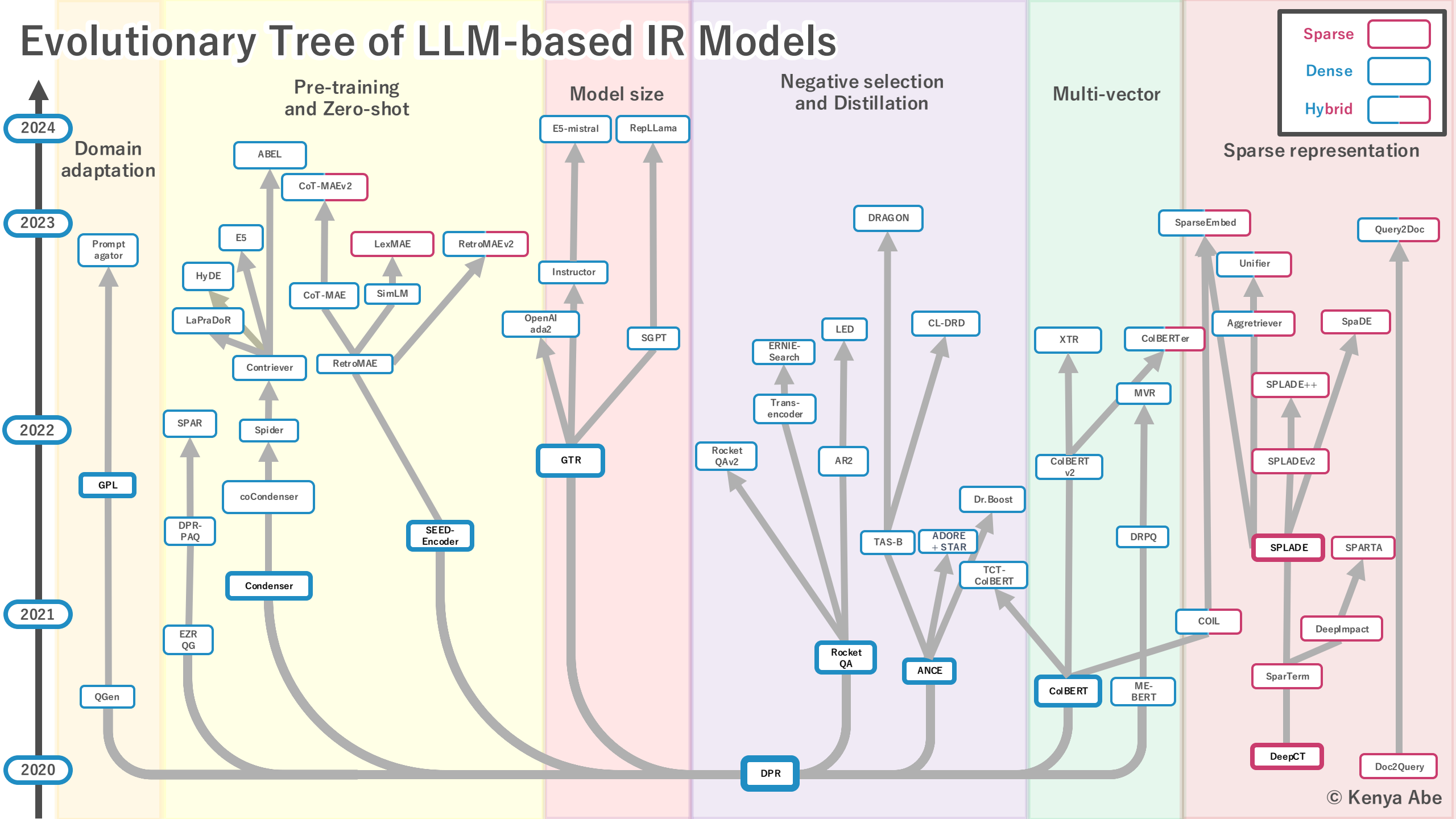
- クエリと文書はBag-of-wordsによって表現され、語彙数次元の疎なベクトルによって表現
- スコアは両ベクトルの内積によって計算
- ☹ クエリ中の語が文書に含まれない場合はスコア0

Karpukhin et al. "Dense passage retrieval for open-domain question answering." EMNLP 2020.



- クエリと文書は検索用に学習された大規模言語モデルによって密なベクトルによって表現
- スコアは両ベクトルの内積によって計算
- ☺ クエリと文書に同じ語がない場合でもベクトルが似ていればスコア大

Evolutionary Tree of LLM-based IR Models

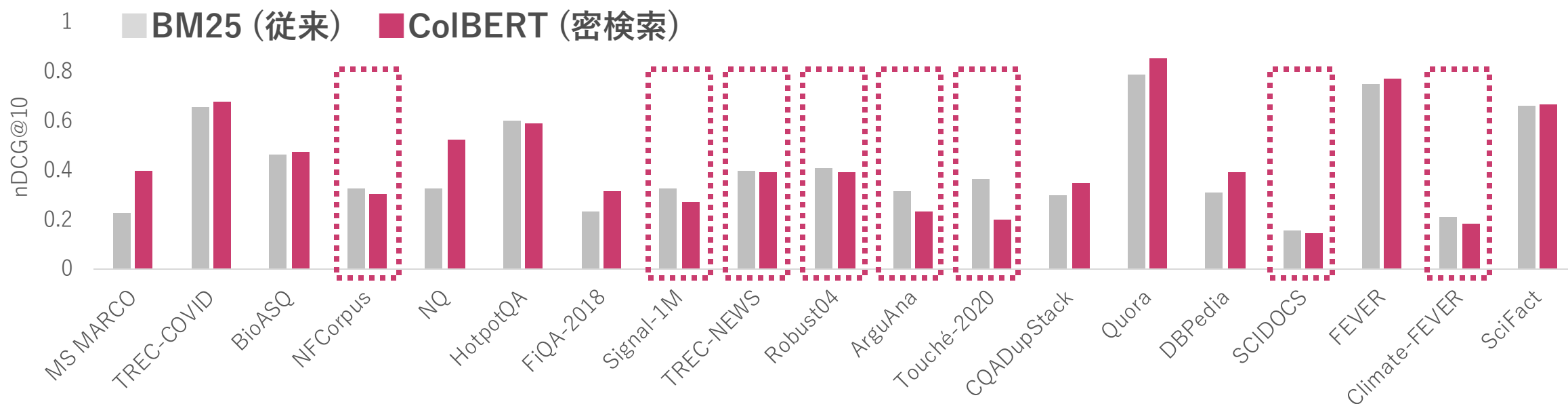


密検索であれば,

- ・ Q. 従来のアルゴリズムよりも常に性能がいい？
- ・ Q. 苦手な検索はない？
- ・ Q. 変わった挙動を示さない？

→ **A. いいえ**

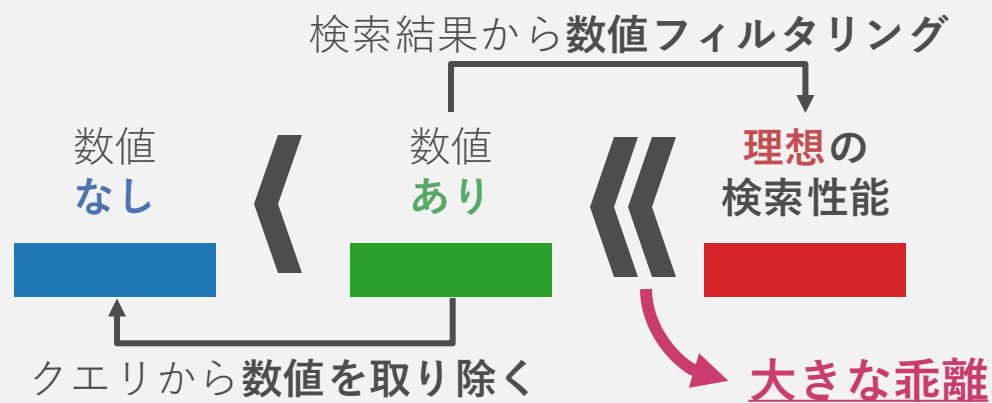
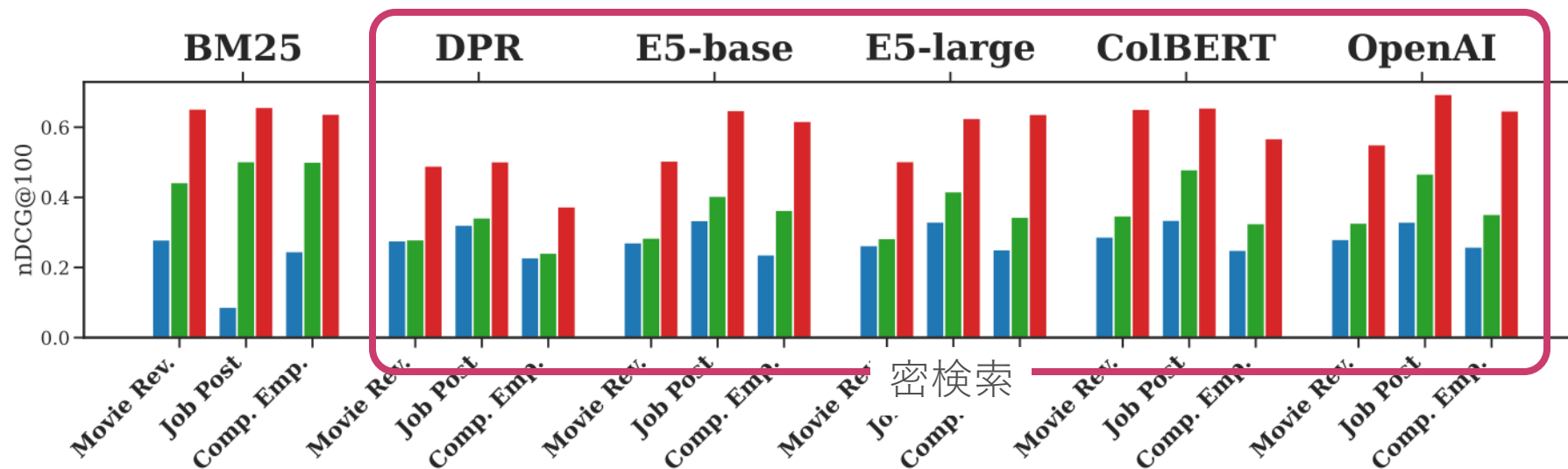
従来の検索アルゴリズムに劣る場合もある



- **Out-of-domain Generalizability** の問題が指摘されている
 - 学習に用いられたドメイン（上記の場合MS MARCO）とは大きく異なるドメインでの性能低下

数値を含む検索クエリの性能が低い

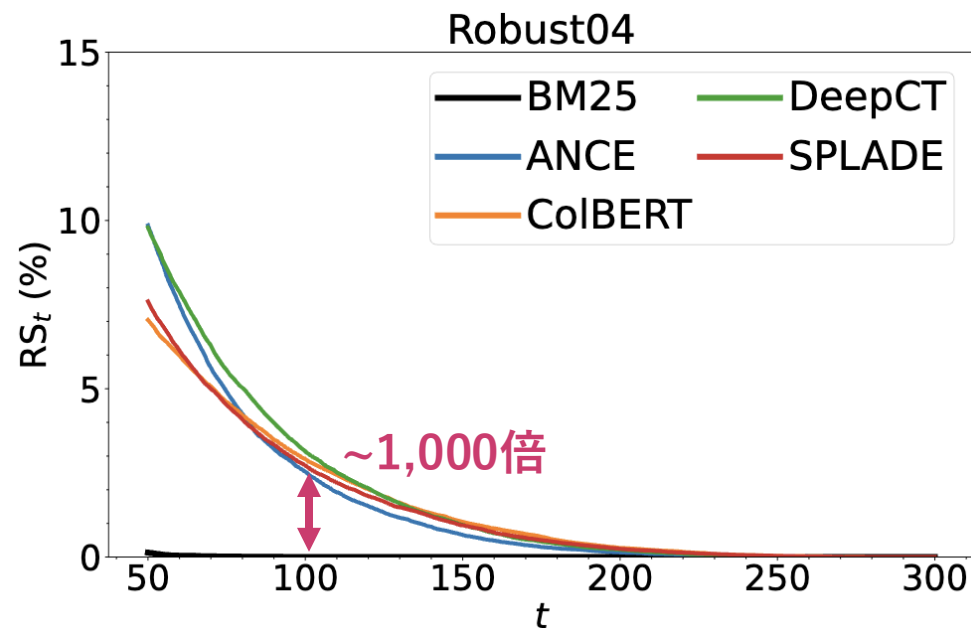
Fujimaki and Kato: Investigating the Performance of Dense Retrievers for Queries with Numerical Conditions. ECIR 2025.



数値を多少理解している
しかし、改善の余地は大きい

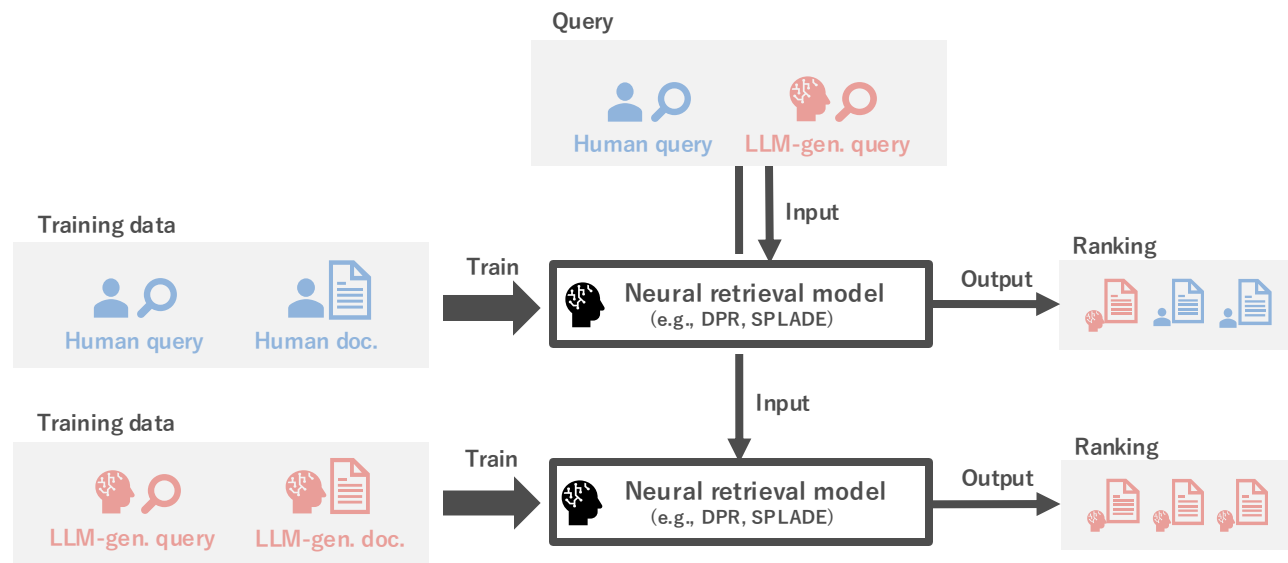
従来の検索アルゴリズムには見られない不都合な挙動を示す

Usuha et al. Over-penalization for Extra Information in Neural IR Models. CIKM 2024.



文書に1文追加したとき，100位以上，その文書の順位を下げてしまう場合（過剰罰則）が従来の2~1,000倍

仲地 and 加藤. 大規模言語モデルによる生成文書・生成クエリを用いた検索モデルの学習と分析. DEIM 2025.



生成AIで生成した文書を密検索用の学習データに含めると，生成AIで生成した文書を優先してしまう傾向（ソースバイアス）を強めてしまう

- 検索システムの顧客は人間から生成AIに
- AI技術による検索システムの改善が進む，が，課題も多い

