

THE FUTURE OF SEARCH BEER & DATA 3

Lucas Farris <lucas@farris.com.br>

WHAT'S GENAI EVEN GOOD FOR?

Over the past few years, I've gotten many questions from business owners regarding how they can use Gen AI to improve their business. One of the best answers is: Retrieval Augmented Generation.

Ex: <https://med.stanford.edu/>

CONTENTS

- Measuring Good Search
- Search Maturity
- Four-Stage Recommender System
- Retrieval Augmented Generation
- Cloud Example



MEASURING GOOD SEARCH


A fancy model is less important than a good evaluation framework




MEASURING

| Accuracy | Relevance | Novelty | Serendipity | Diversity |
|--|---|---|---|--|
| <p>Always remember the precision-recall trade-off. Measuring nDCG is always a good idea.</p> | <p>Human relevance (judgement) vs tracking relevance (presentation bias).</p> | <p>Users are likely to engage with new items if they are relevant</p> | <p>Non "obvious" recommendations make users happier</p> | <p>There should be enough variability between search results</p> |


amazon.pl Zaloguj się  

potrzebuje więcej energia 


Dostawa do: 50073 – Aktualizacja lokalizacji  [Dołącz do Prime](#)

prime Zabawki Elektronika Ogród [Filtry](#)

[Dowiedz się więcej o tych wynikach.](#)

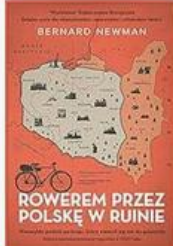


Miernik kosztów energii zużycie energii elektrycznej potrzebuje, 0,000 kh-999kwhh Connector Miernik Intelligentny monitor energii cyfrowej z wyświetlacze...


3,9  (5)

78⁶⁷ zł (491,69zł/m)

Otrzymaj w dniach **sobota, 27 lipca** – **środa, 7 sierpnia**
DARMOWA dostawa

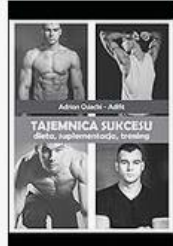


ROWEREM PRZEZ POLSKĘ W RUINIE
Newman Bernard

5,0  (1)

Oprawa miękka
33⁵² zł


DARMOWA dostawa przez Amazon
Wysyłka zazwyczaj w ciągu 1–3 miesięcy




Tajemnica Sukcesu - Dieta, trening i suplementacja
Adrian Osiecki

Oprawa miękka
36⁷⁵ zł

Dostawa do dnia: **środa, 26 czerwca**
DARMOWA dostawa przez Amazon



AntHouse. es - Super pokarm dla mrówek czerwony nektar żywność 50 ml | Idealny dodatek do terrarium lub mrowiska, pokarm dla mrówek, aby mieć energię,...

4,0  (18)

43⁷⁵ zł

Otrzymaj w dniach **poniedziałek, 1 lipca** – **sobota, 6 lipca**
45,36 zł za dostawę

[Dostępne sztuki: 8](#) [zamów teraz](#)



WHAT MAKES SEARCH HARD

Dynamic Inventory

- Maybe your products are only sold a limited number of times
- Maybe they sell-out fast (e.g., luxury cars, jewelry, houses)
- Is everything you learn about them lost?

One-Time Users

- A common problem of dating sites, car manufacturers, and real-estate
- Some businesses have users that buy once come back only years later (or never)

Scarce Data

- Maybe your products have a small amount of features
- Maybe your users are mostly anonymous
- Maybe your data is in a CRM software you can't access
- Maybe you're not measuring

SEARCH MATURITY

A two-axes framework to evaluate and advance
a search experience

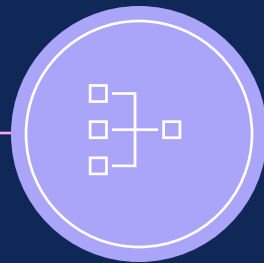


LANGUAGE MATURITY



Level 1

Keyword Matching:
Inventory is indexed (inverse index), filters and sorting perform exact matches against attributes



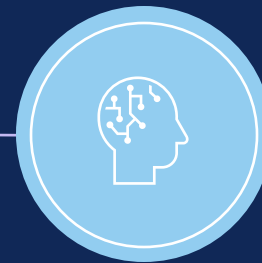
Level 2

Taxonomies:
Items are catalogued and grouped under entities, cross-entity categorizations are built (ontologies), synonym dictionaries are built in the search



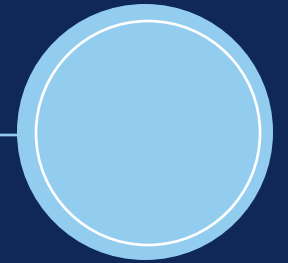
Level 3

Query Intent:
Semantics, query rewrites (i.e., did you mean...). System knows if you're looking for an article, for a product, or something else.



Level 4

Knowledge Graphs:
Knowledge beyond structured data (e.g., images, audio, videos) connecting them to text



airbnb Mediterranean Sea Any week Add guests Airbnb your home

Your search Rooms Countryside **Amazing pools** Cabins Lake Play Farms Tiny homes Vineyards Design Amazing views National parks Skiing Trending Filters

Over 1,000 amazing pools

Guest favorite

Home in Ostuni, Italy ★ 4.93 (257)
1 SETTE CONI - TRULLO EDERA
1 double bed
Sep 14 - 20
1,148 zł night

Guest favorite

Farm stay in Terricciola, Italy ★ 5.0 (22)
Tuscany | Farmhouse with pool and restaurant
3 beds
Sep 7 - 12
919 zł 831 zł night

Guest favorite

Villa in Provincia di Lecce, Italy ★ 4.92 (13)
Masseria Quadrelli: an oasis in the heart of Salento
20 beds
Jul 15 - 20
7,567 zł night

RANKING MATURITY



Level 1

Term-Frequency
Relevance is measured based on how many times the search keywords appear in documents (i.e., a product's description). Commonly used algorithms are TF-IDF and BM25.



Level 2

Collaborative Filtering
You track the rate between certain web events (i.e., page views) and rank items based on how they perform. Usual metrics for this are Click-Through Rate (CTR) or Conversion Rate (CR).



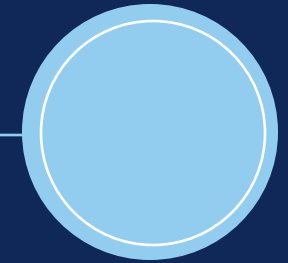
Level 3

Model-Based
You use your inventory attributes and performance to train a regression based to predict how each item will perform. Usually non-negative matrix factorization (NNMF) and XGBoost are good choices here.



Level 4

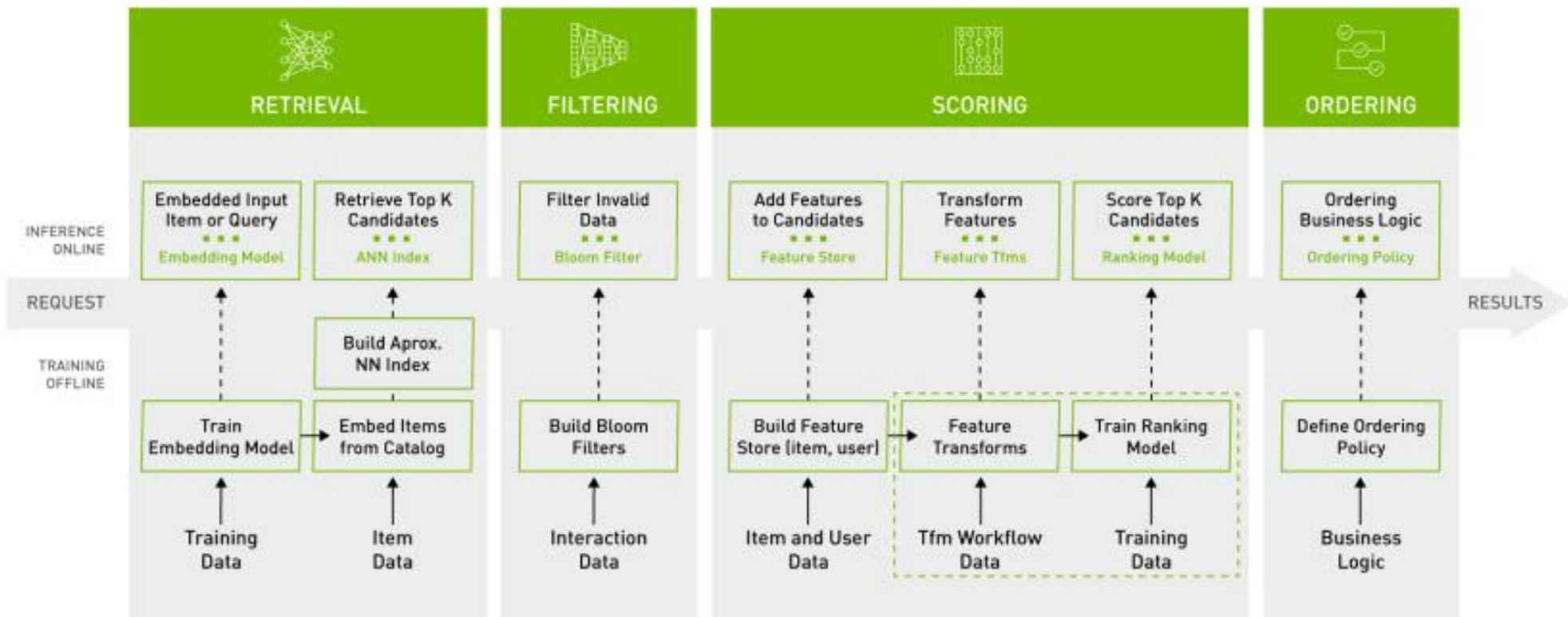
Personalization
The relevance model is trained with user features and unstructured data, and inference happens in real time. Usually in this case we would need deep neural networks and GPUs.



SOLUTION ARCHITECTURE

The four-stage recommender system





WORD EMBEDDINGS

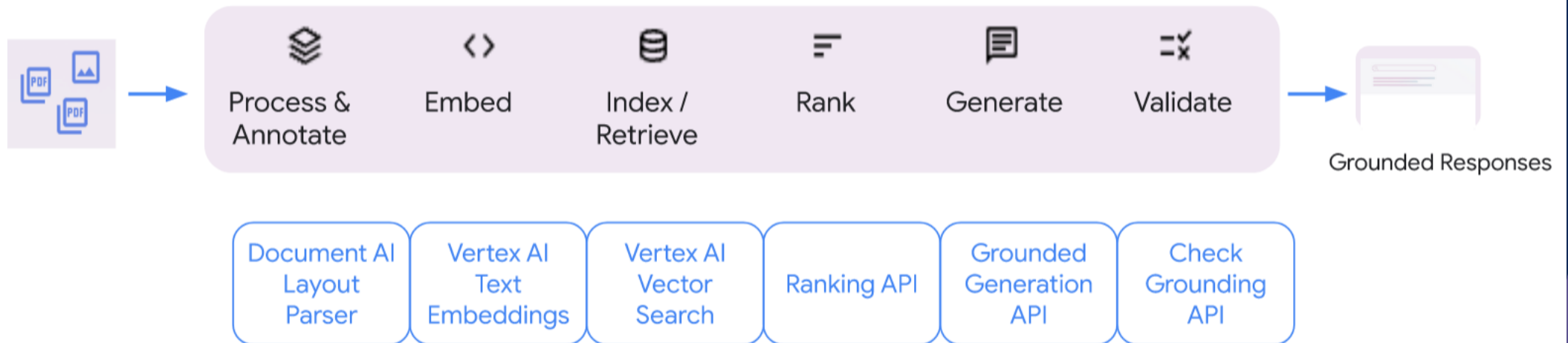
| Alternative Labels | Synonyms | Taxonomy | Ontology | Knowledge Graph |
|--|---|---|---|---|
| <p>Words, acronyms, or expressions have the exact same meaning</p> | <p>Tokens with very similar meaning</p> | <p>Relationships between tokens and their categories or classes</p> | <p>Relationships between tokens, for instance in terms of how they interact</p> | <p>Mappings and relationships between entities and their related concepts</p> |

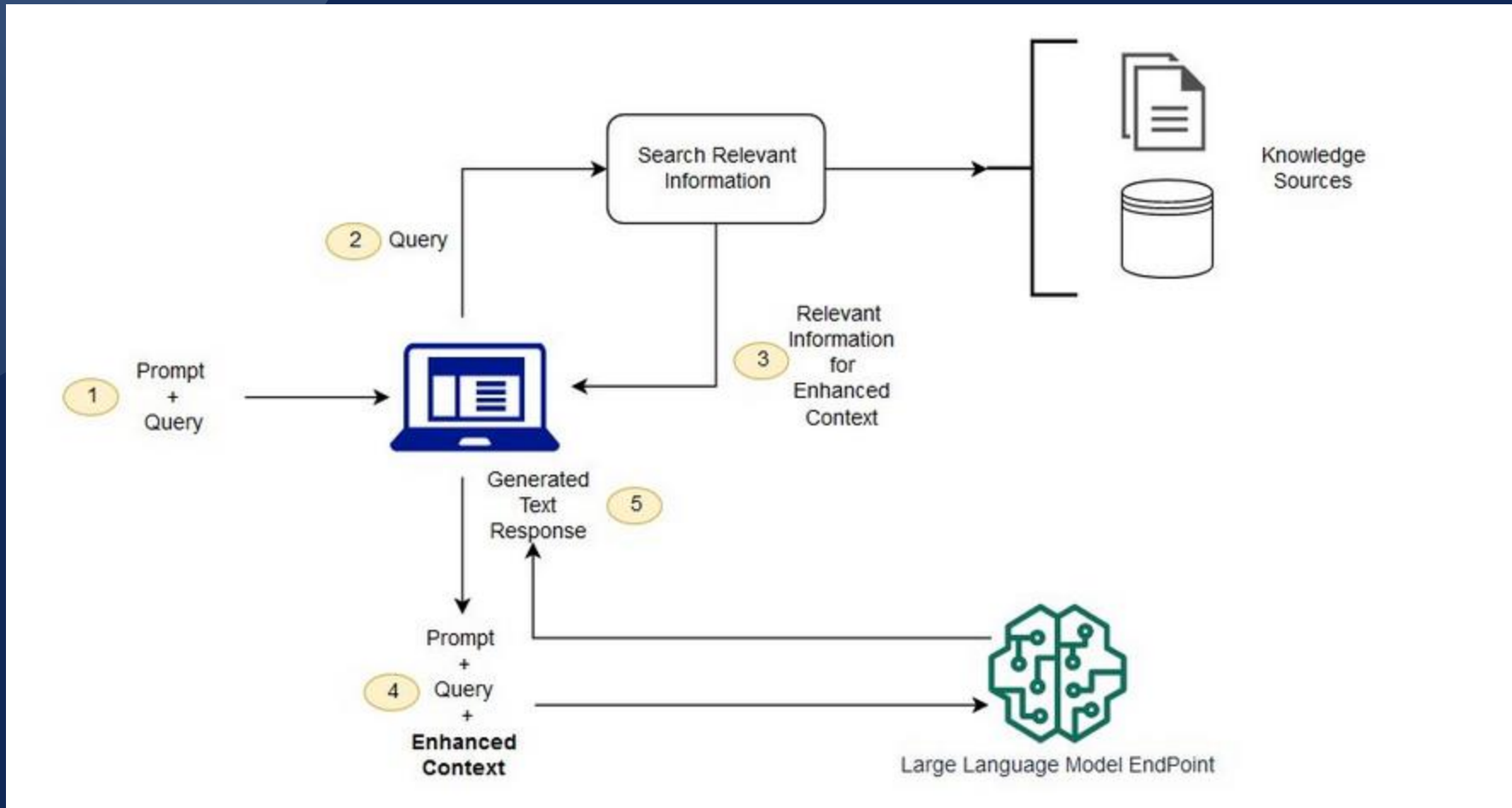
RETRIEVAL AUGMENTED GENERATION

What LLMs can do for search



Vertex AI APIs for RAG





DEMO TIME

Google Cloud Platform Example



THANK YOU

Lucas Farris

lucas@farris.com.br

