elastic | The Search Analytics Company

Introduction to RAG

Anastasiia Havriushenko

ML Engineer

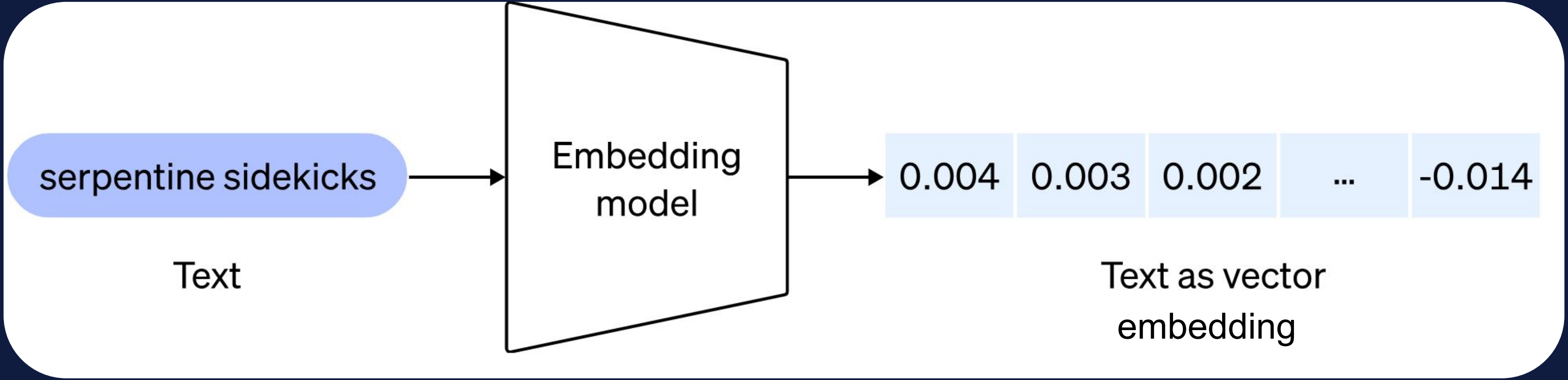
Keep in touch

Overview

1. Text embeddings
2. LLM basics
3. RAG architecture
4. Demo

elastic

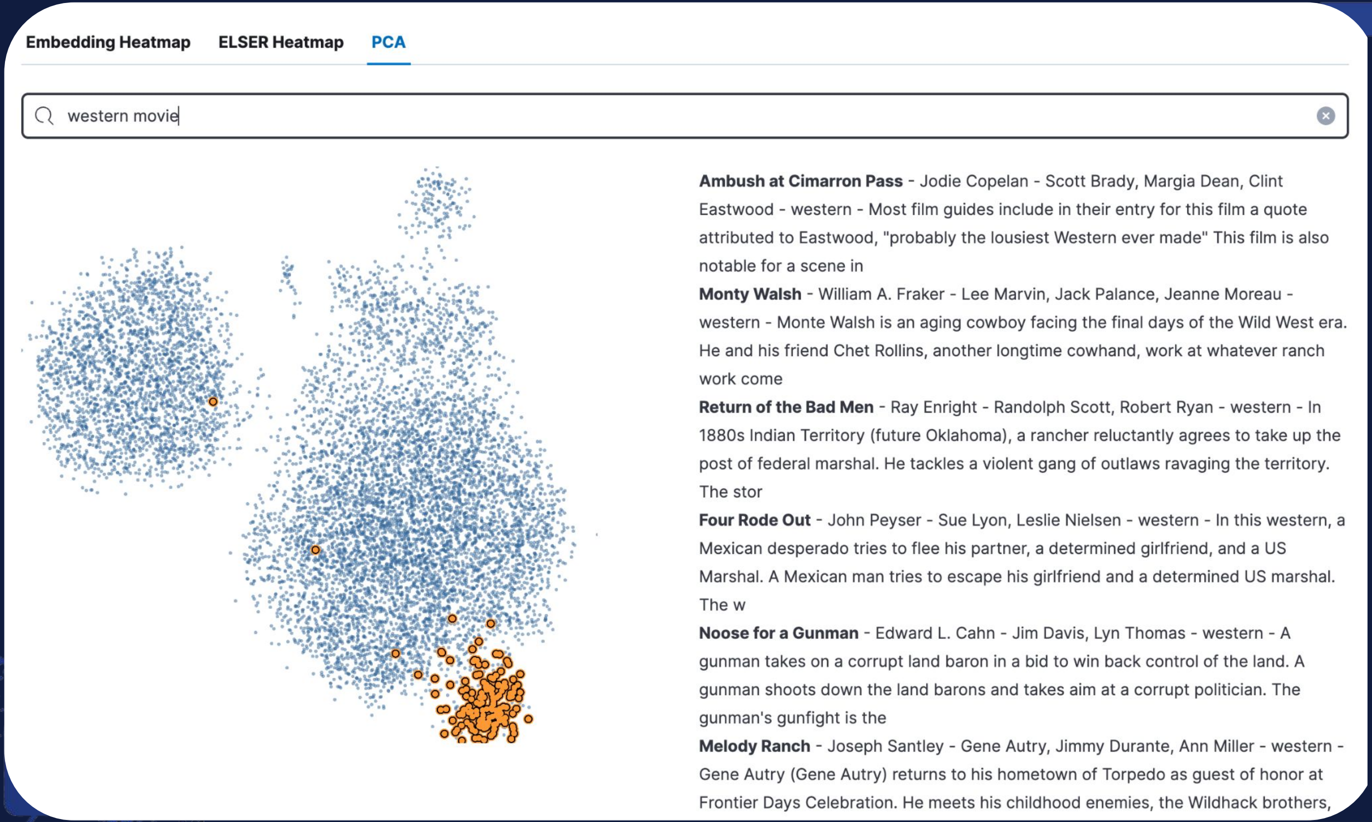
Vector embeddings



Python examples: word2vec, text2vec

Image: OpenAI

Embedding visualization



western movie

Ambush at Cimarron Pass - Jodie Copelan - Scott Brady, Margia Dean, Clint Eastwood - western - Most film guides include in their entry for this film a quote attributed to Eastwood, "probably the lousiest Western ever made" This film is also notable for a scene in

Monty Walsh - William A. Fraker - Lee Marvin, Jack Palance, Jeanne Moreau - western - Monte Walsh is an aging cowboy facing the final days of the Wild West era. He and his friend Chet Rollins, another longtime cowhand, work at whatever ranch work come

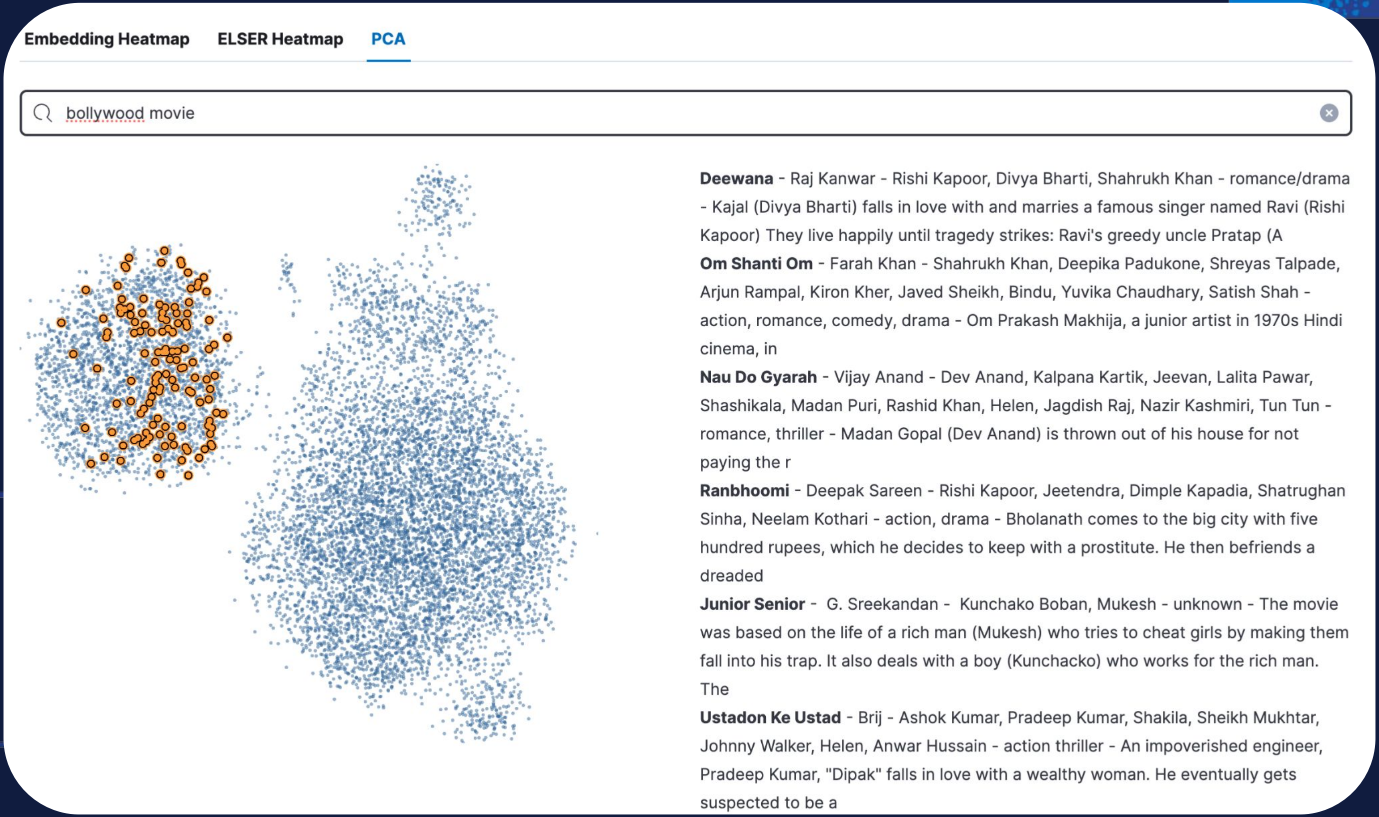
Return of the Bad Men - Ray Enright - Randolph Scott, Robert Ryan - western - In 1880s Indian Territory (future Oklahoma), a rancher reluctantly agrees to take up the post of federal marshal. He tackles a violent gang of outlaws ravaging the territory. The stor

Four Rode Out - John Peyser - Sue Lyon, Leslie Nielsen - western - In this western, a Mexican desperado tries to flee his partner, a determined girlfriend, and a US Marshal. A Mexican man tries to escape his girlfriend and a determined US marshal. The w

Noose for a Gunman - Edward L. Cahn - Jim Davis, Lyn Thomas - western - A gunman takes on a corrupt land baron in a bid to win back control of the land. A gunman shoots down the land barons and takes aim at a corrupt politician. The gunman's gunfight is the

Melody Ranch - Joseph Santley - Gene Autry, Jimmy Durante, Ann Miller - western - Gene Autry (Gene Autry) returns to his hometown of Torpedo as guest of honor at Frontier Days Celebration. He meets his childhood enemies, the Wildhack brothers,

Embedding visualization



bollywood movie

Deewana - Raj Kanwar - Rishi Kapoor, Divya Bharti, Shahrukh Khan - romance/drama - Kajal (Divya Bharti) falls in love with and marries a famous singer named Ravi (Rishi Kapoor) They live happily until tragedy strikes: Ravi's greedy uncle Pratap (A

Om Shanti Om - Farah Khan - Shahrukh Khan, Deepika Padukone, Shreyas Talpade, Arjun Rampal, Kiron Kher, Javed Sheikh, Bindu, Yuvika Chaudhary, Satish Shah - action, romance, comedy, drama - Om Prakash Makhija, a junior artist in 1970s Hindi cinema, in

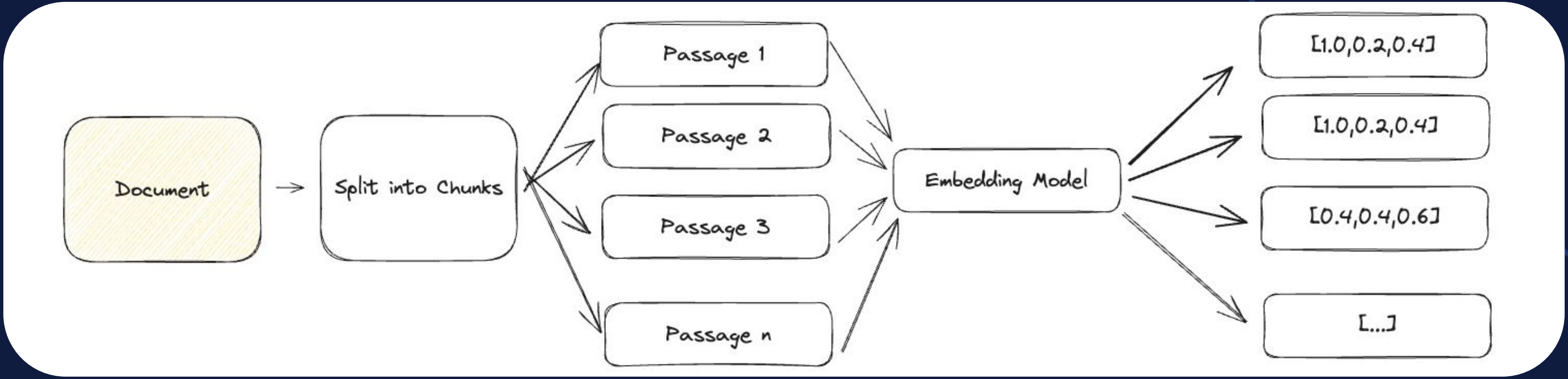
Nau Do Gyarah - Vijay Anand - Dev Anand, Kalpana Kartik, Jeevan, Lalita Pawar, Shashikala, Madan Puri, Rashid Khan, Helen, Jagdish Raj, Nazir Kashmiri, Tun Tun - romance, thriller - Madan Gopal (Dev Anand) is thrown out of his house for not paying the r

Ranbhoomi - Deepak Sareen - Rishi Kapoor, Jeetendra, Dimple Kapadia, Shatrughan Sinha, Neelam Kothari - action, drama - Bholanath comes to the big city with five hundred rupees, which he decides to keep with a prostitute. He then befriends a dreaded

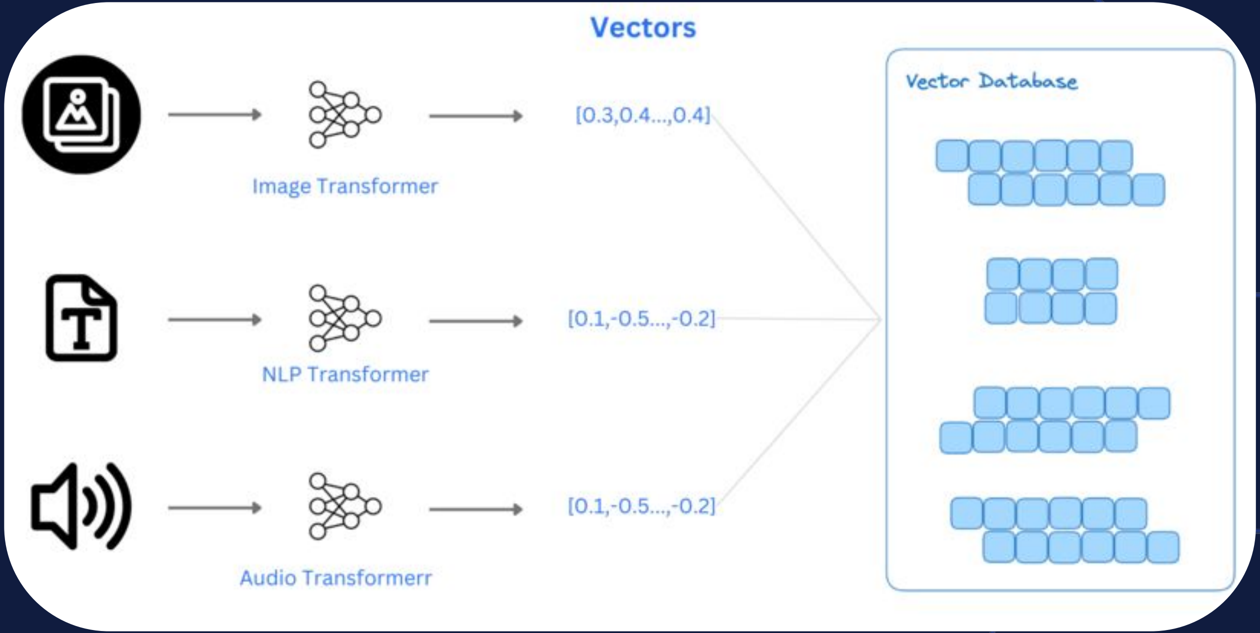
Junior Senior - G. Sreekandan - Kunchako Boban, Mukesh - unknown - The movie was based on the life of a rich man (Mukesh) who tries to cheat girls by making them fall into his trap. It also deals with a boy (Kunchacko) who works for the rich man. The

Ustadon Ke Ustad - Brij - Ashok Kumar, Pradeep Kumar, Shakila, Sheikh Mukhtar, Johnny Walker, Helen, Anwar Hussain - action thriller - An impoverished engineer, Pradeep Kumar, "Dipak" falls in love with a wealthy woman. He eventually gets suspected to be a

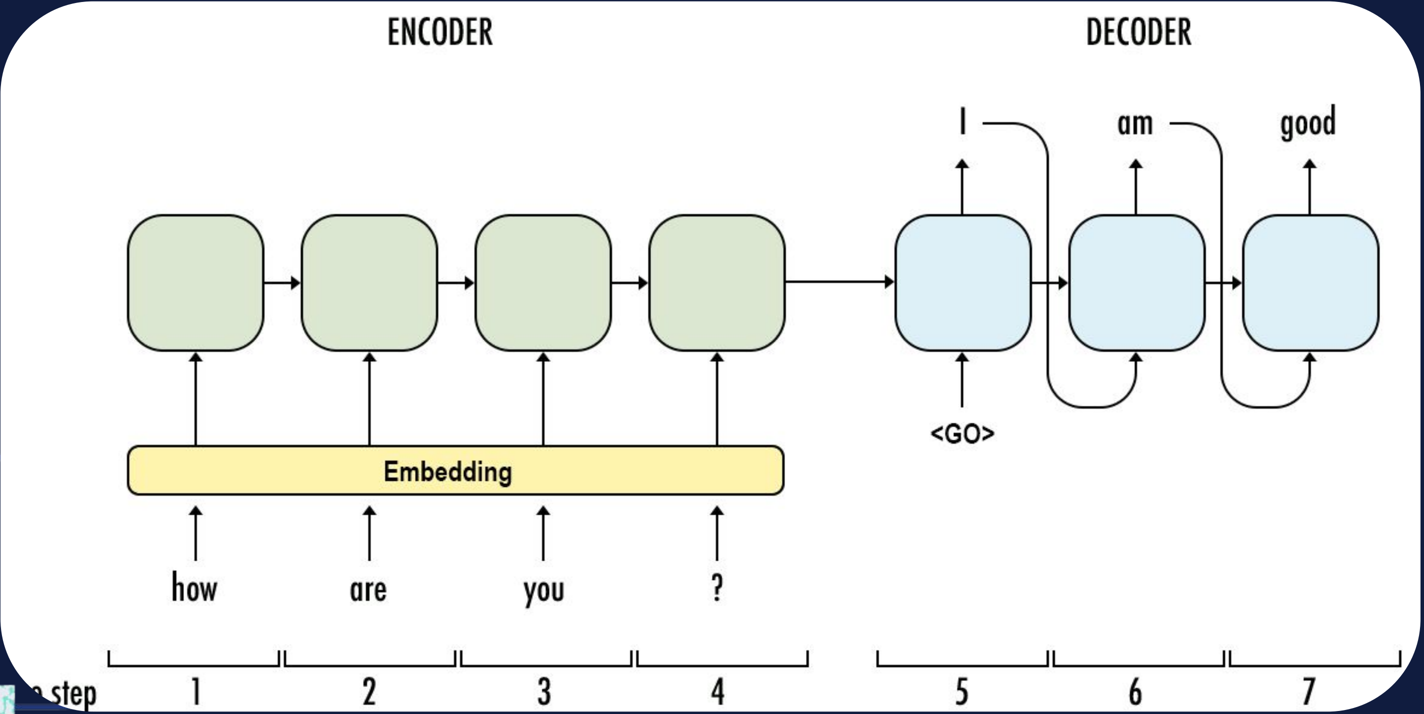
Vector embeddings



Vector embeddings



NLP design


ENCODER
DECODER
I
am
good
Embedding
<GO>
how
are
you
?
step
1
2
3
4
5
6
7

Evalution of Chatbots

APPLE
Siri
The Coolheaded Celebrity
No one shuts down your sass like Siri.

AMAZON
Alexa
The Housekeeper That Knows Too Much
She knows all about the 'adult novelty items' in your Amazon shopping cart.

GOOGLE
Now
The Proactive Stalker
Just because you know my entire browser history doesn't make us friends.

MICROSOFT
Tay
The Overly Impressionable Millennial
Maybe some therapy will sort out her psychotic algorithm issues.

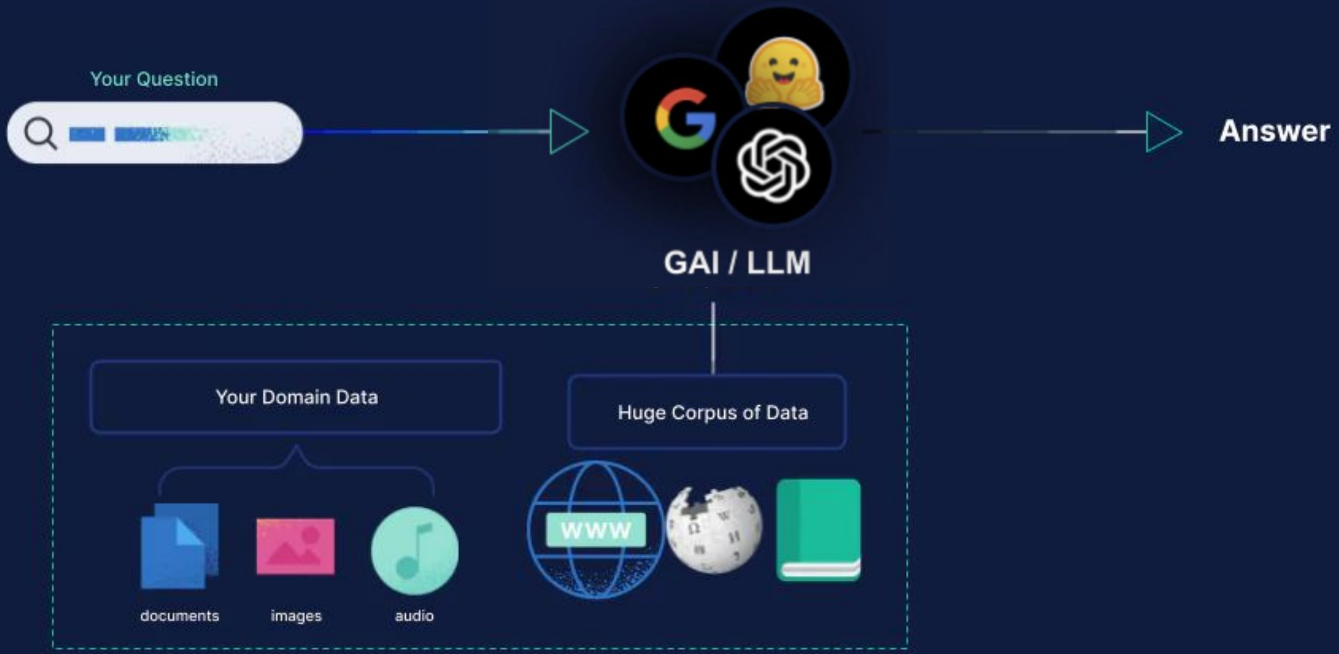
MICROSOFT
Cortana
The Perceptive Friend
Inexplicably has their shit together more than you.

FACEBOOK
M
Life's Call Centre
Answer times may vary.

Common LLM challenges:

- Hallucinations
- Lack of domain knowledge
- Frozen parametric knowledge
- Privacy concerns
- Expensive to (re)train

Domain Specific Pre-Training for LLMs



Your Question

GAI / LLM

Answer

Your Domain Data

documents

images

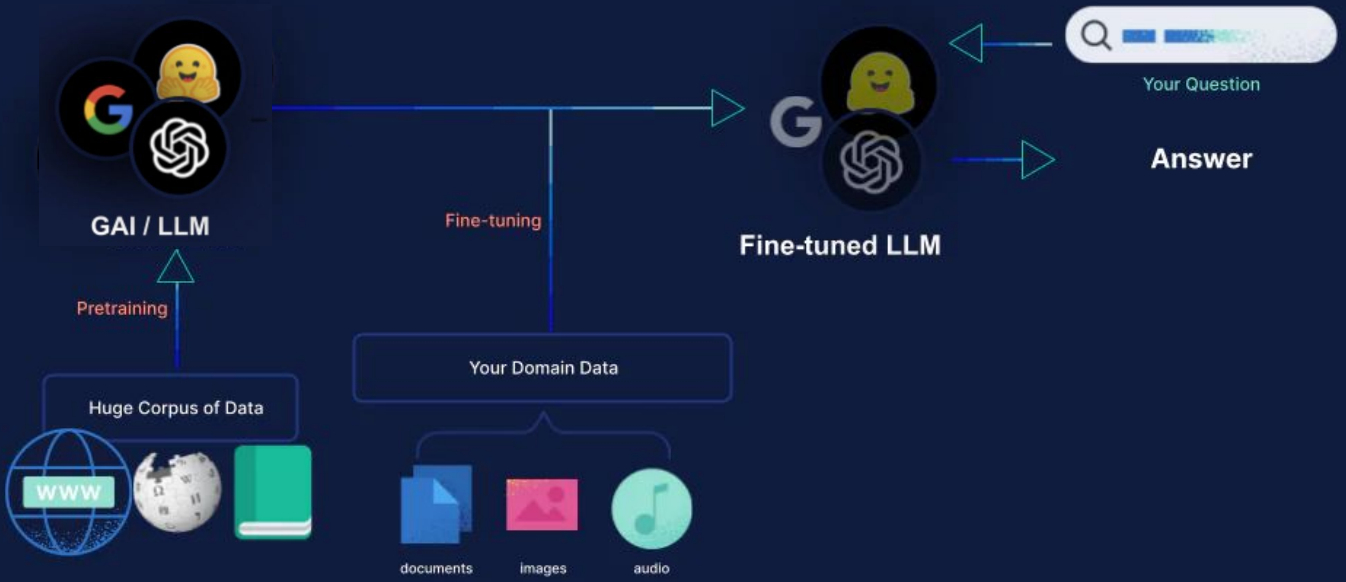
audio

Huge Corpus of Data

More info

elastic

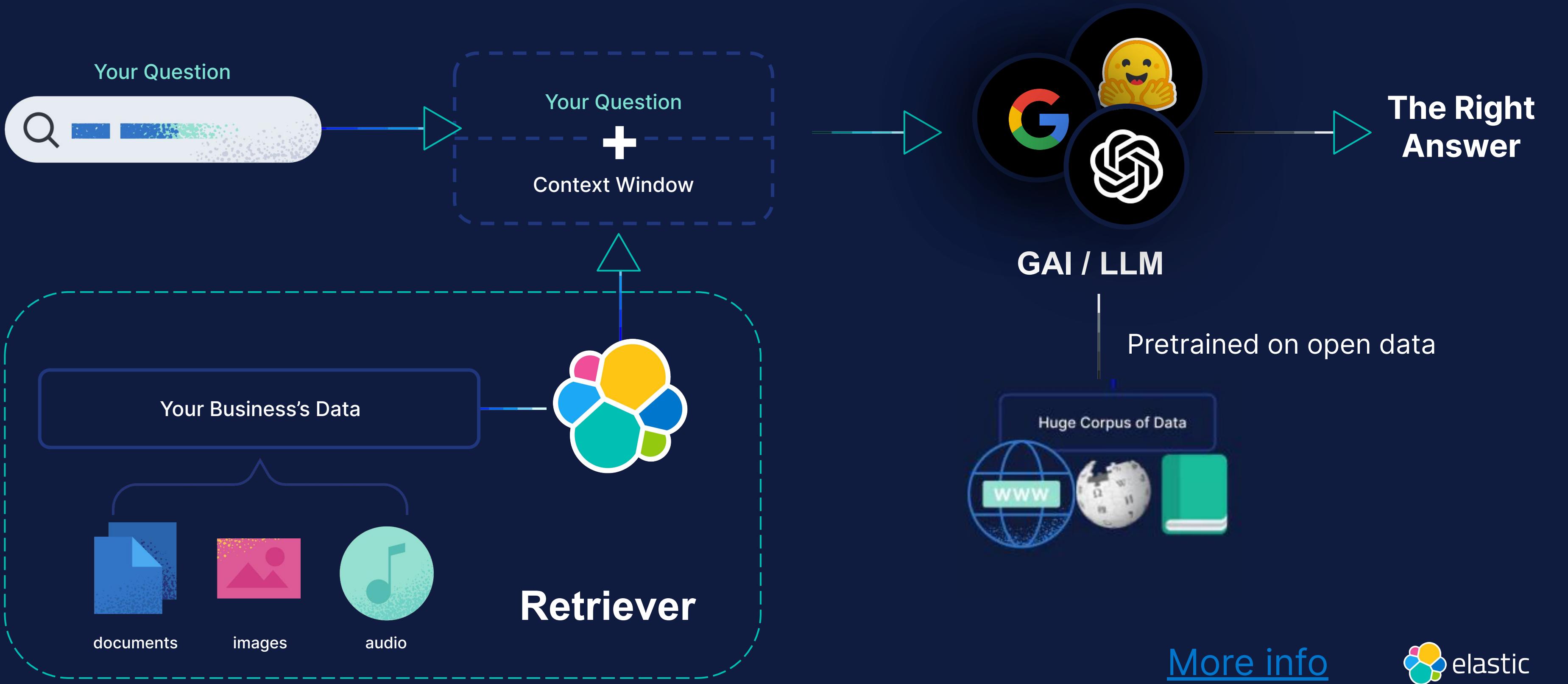
Domain Specific Fine-Tuning for LLMs



More info

elastic

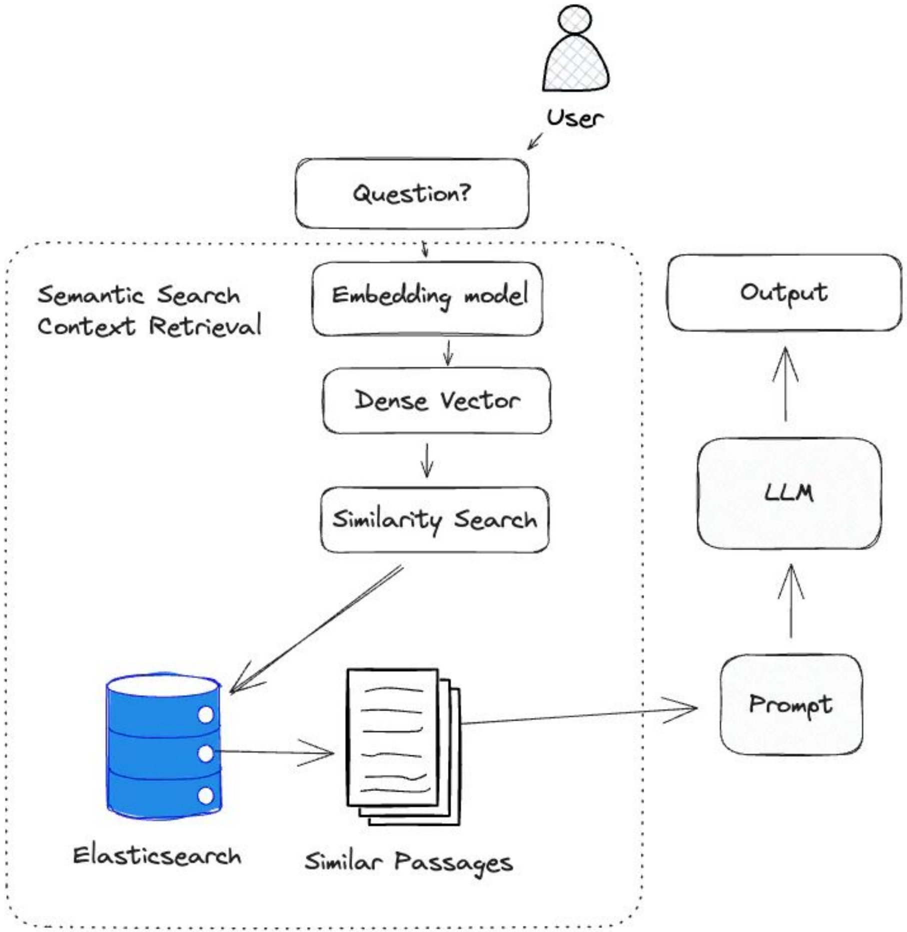
RAG Elastic workflow



More info

elastic

RAG architecture



Demo

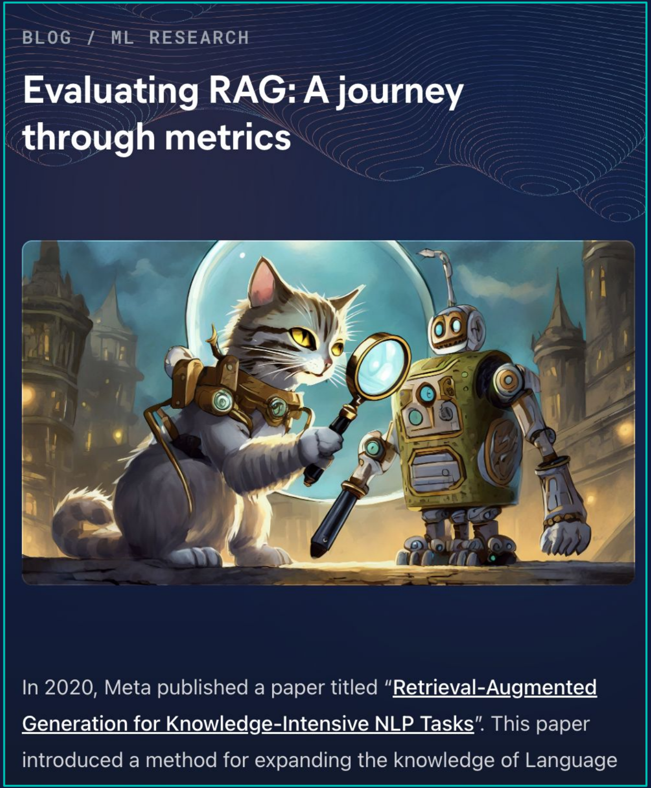
Link code
Link app

Resources for developers

New!

elastic.co/search-labs

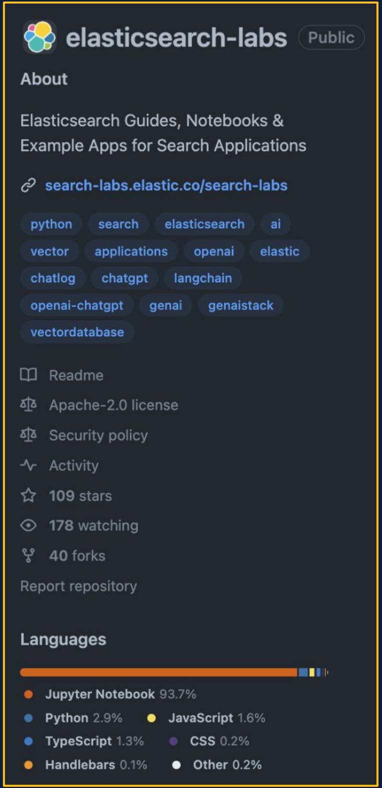
github.com/elastic/elasticsearch-labs



BLOG / ML RESEARCH

Evaluating RAG: A journey through metrics

In 2020, Meta published a paper titled "Retrieval-Augmented Generation for Knowledge-Intensive NLP Tasks". This paper introduced a method for expanding the knowledge of Language



elasticsearch-labs Public

About

Elasticsearch Guides, Notebooks & Example Apps for Search Applications

search-labs.elastic.co/search-labs

python search elasticsearch ai vector applications openai elastic chatlog chatgpt langchain openai-chatgpt genai genaistack vectordatabase

Readme

Apache-2.0 license

Security policy

Activity

109 stars

178 watching

40 forks

Report repository

Languages

● Jupyter Notebook 93.7%
● Python 2.9% ● JavaScript 1.6%
● TypeScript 1.3% ● CSS 0.2%
● Handlebars 0.1% ● Other 0.2%

Generative AI
ML Research
Vector Search
How-Tos
Integrations
Lucene

elastic

Thank you!
Questions?

elastic