elastic

Search, Observability, Security:
a new era

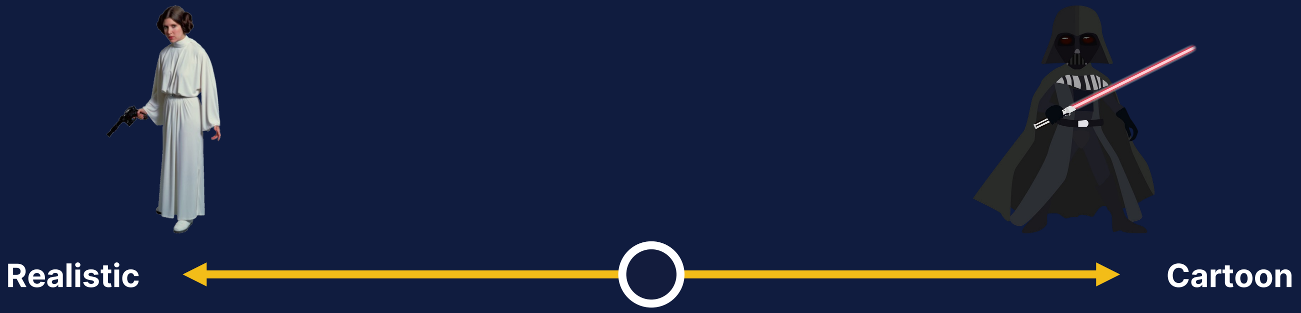
David Pilato | @dadoonet

Elasticsearch

You Know, for **Vector** Search

Embeddings represent your data

Example: 1-dimensional vector



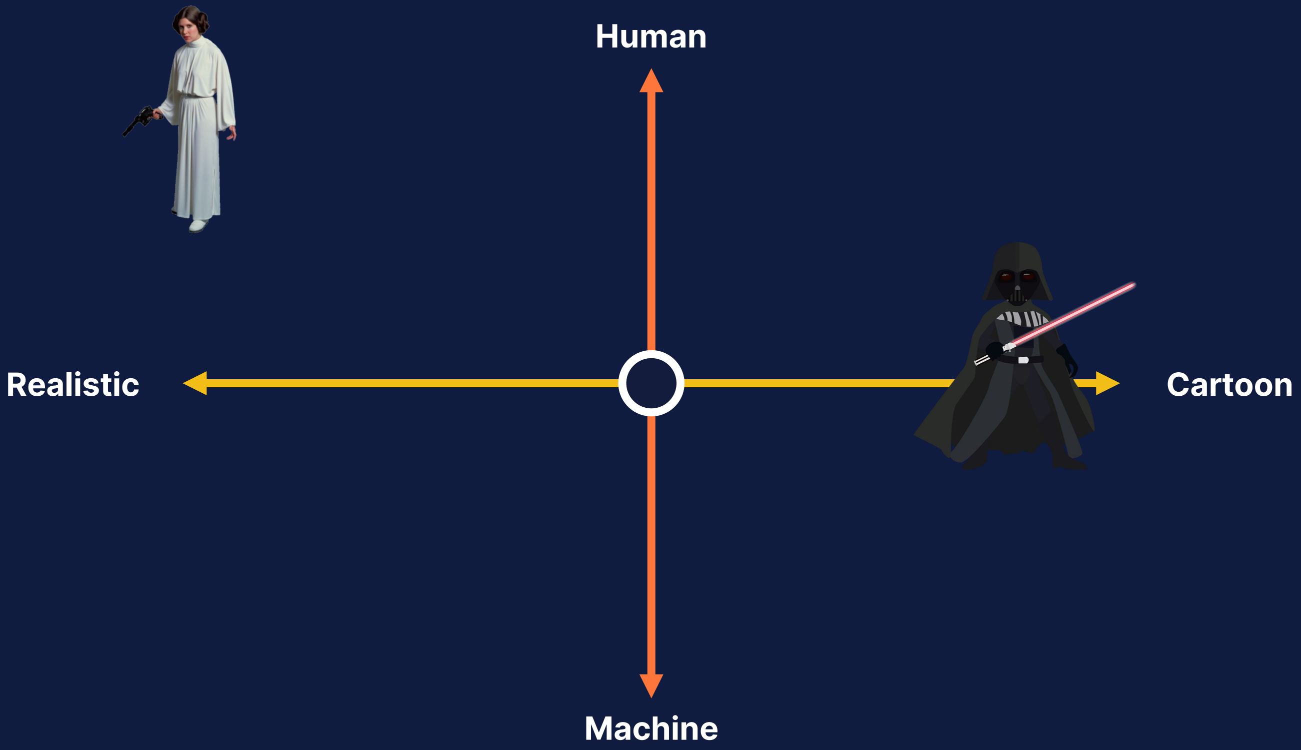
Realistic

Cartoon

Character	Vector
	[-1]
	[1]

Multiple dimensions

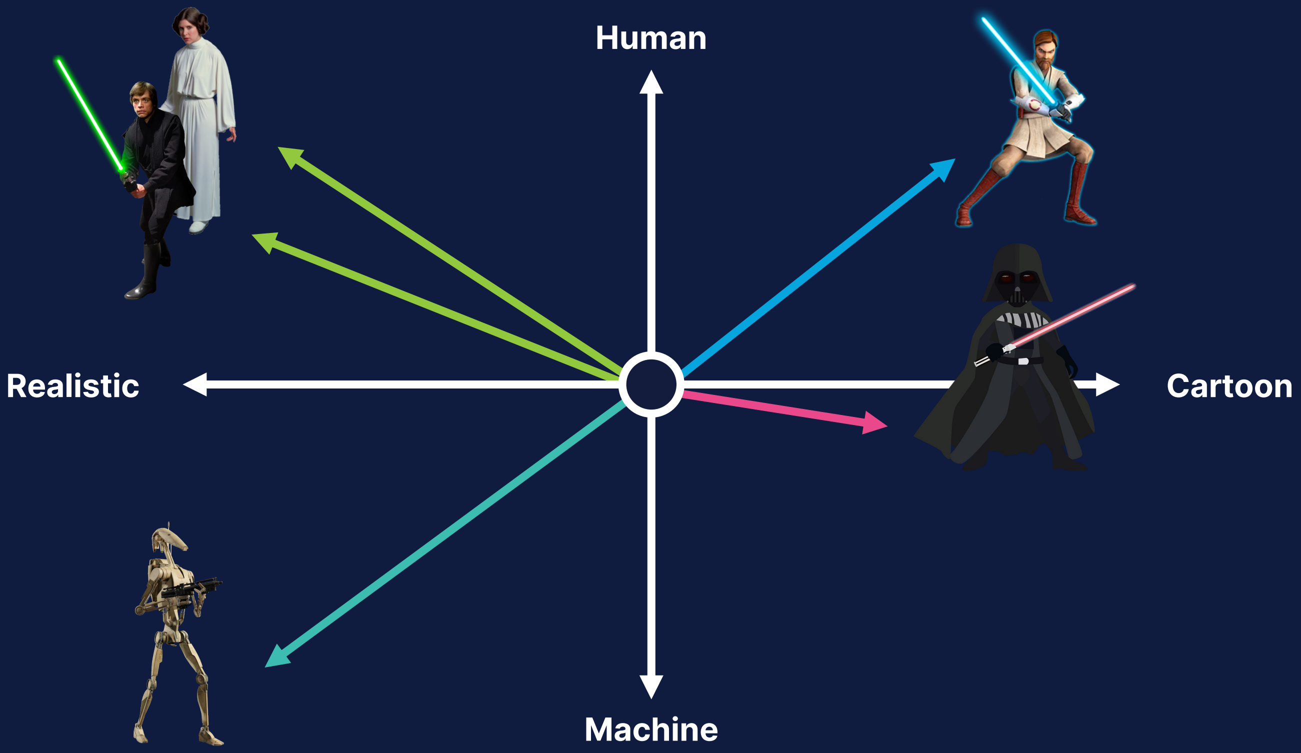
represent different data aspects



Character	Vector
	[-1, 1]
	[1, 0]

Similar data

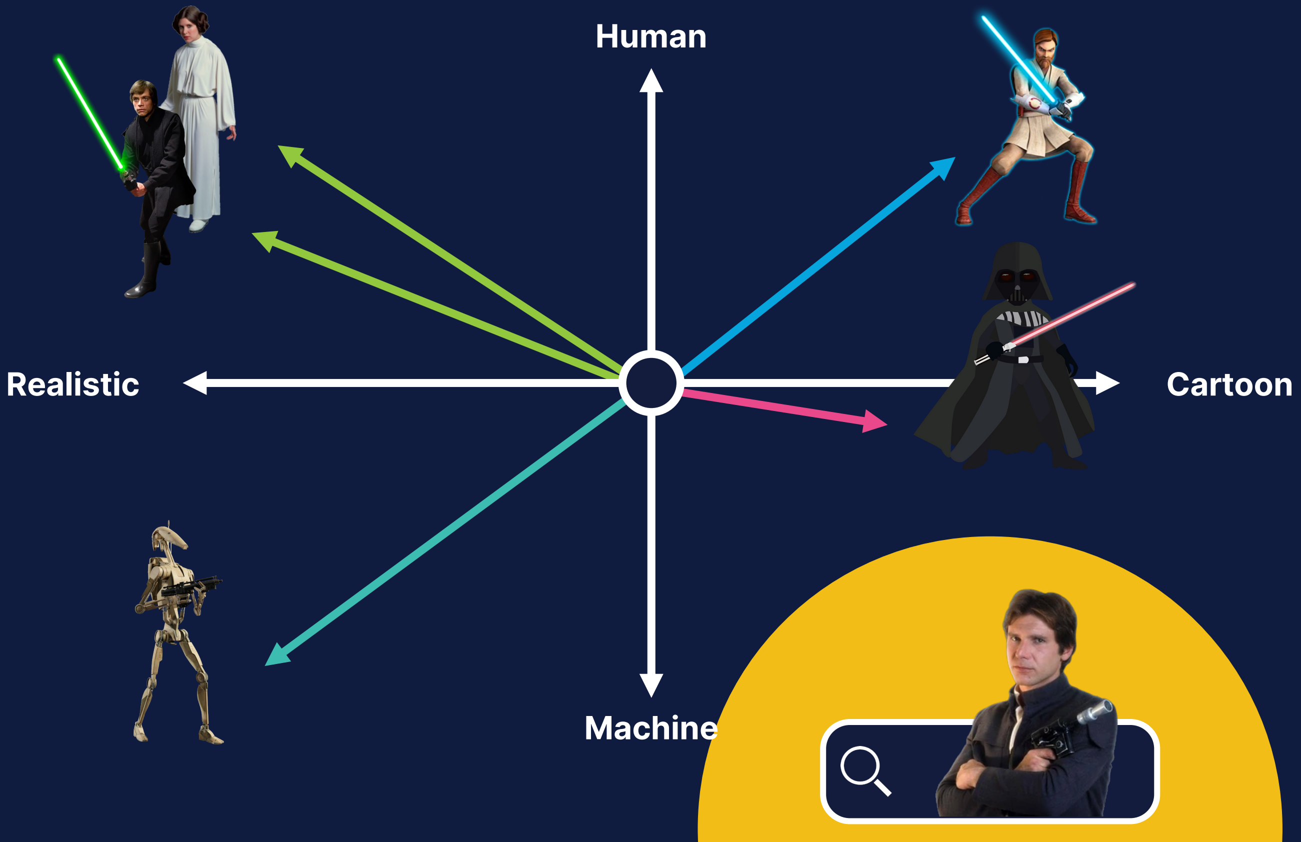
is grouped together



Character	Vector
	[-1.0, 1.0]
	[1.0, 0.0]
	[-1.0, 0.8]

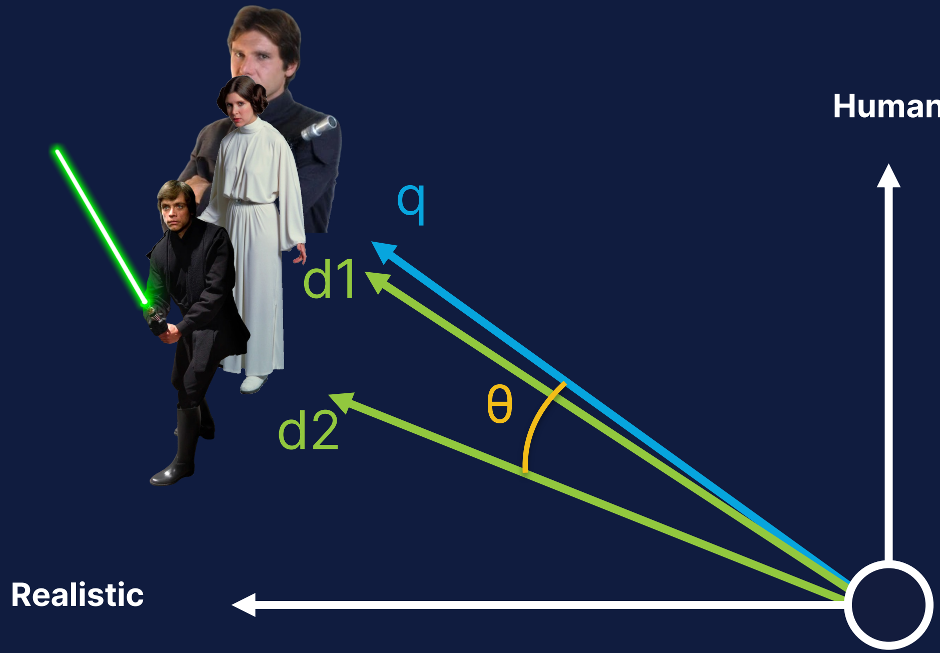
Vector search ranks objects

by similarity (~relevance) to the query



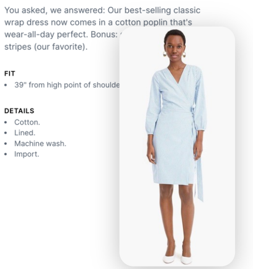
Rank	Result
Query	
1	
2	
3	
4	
5	

Similarity: cosine (cosine)



$$\_score = \frac{1 + cos(\theta)}{2}$$

Data Ingestion and Embedding Generation



You asked, we answered: Our best-selling classic wrap dress now comes in a cotton poplin that's wear-all-day perfect. Bonus:

stripes (our favorite).

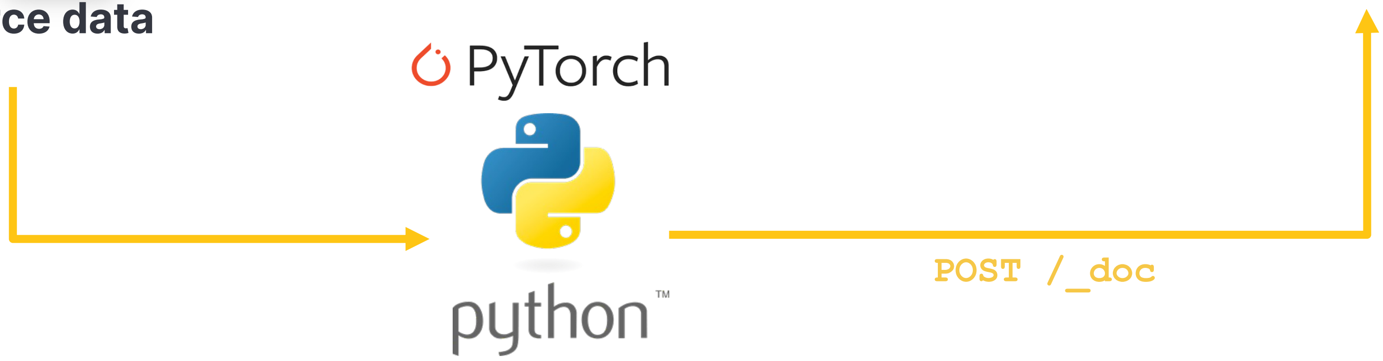
FIT

- 39" from high point of shoulde

DETAILS

- Cotton.
- Lined.
- Machine wash.
- Import.

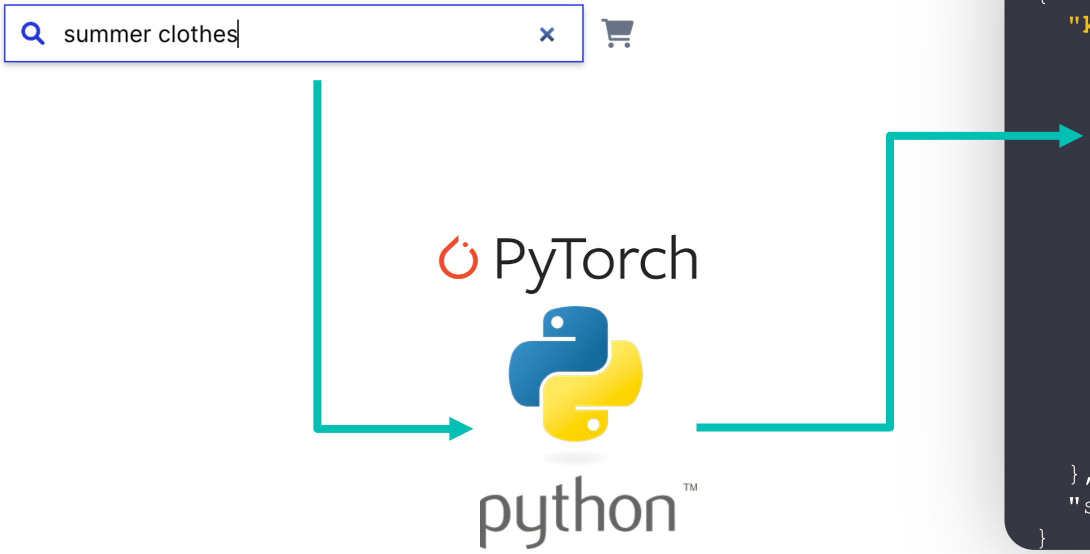
Source data



POST /_doc

```
{
  "_id":"product-1234",
  "product_name":"Summer Dress",
  "description":"Our best-selling…",
  "Price": 118,
  "color":"blue",
  "fabric":"cotton",
  "desc_embedding":[0.452,0.3242,…],
  "img_embedding":[0.012,0.0,…]
}
```

Vector Query



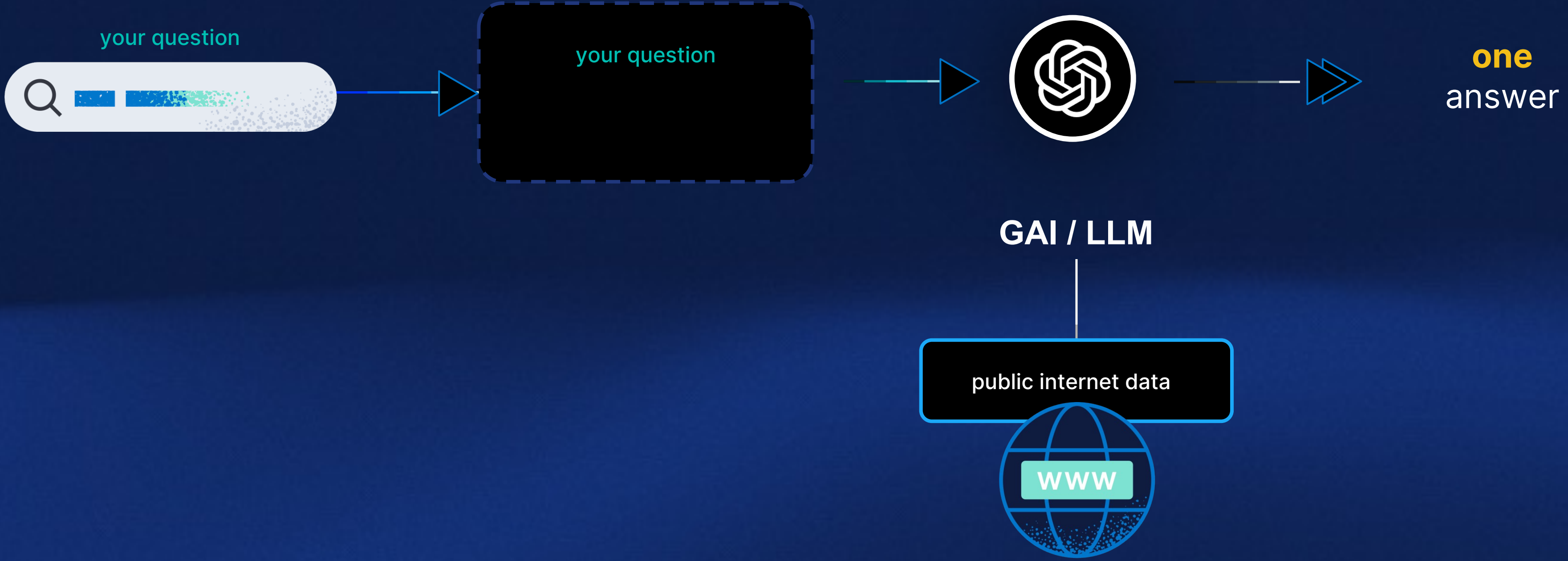
```
GET product-catalog/_search
{
  "knn": {
    "field": "desc_embbeding",
    "k": 5,
    "num_candidates": 50,
    "query_vector": [0.123, 0.244,...],



    "filter": {
       "term": {
         "department": "women"
       }
    }
  },
  "size": 10
}
```

ChatGPT

Elastic and LLM

Gen AI
Search engines

LLM: opportunities and limits



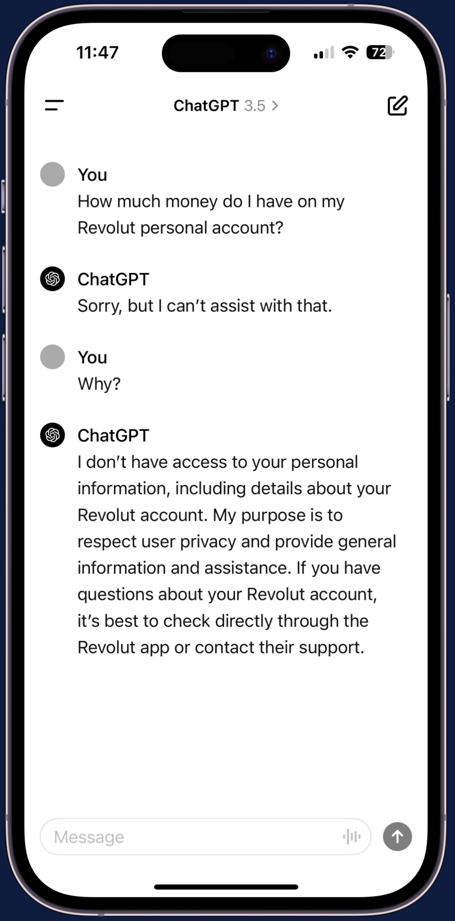
your question

your question

GAI / LLM

public internet data

one
answer



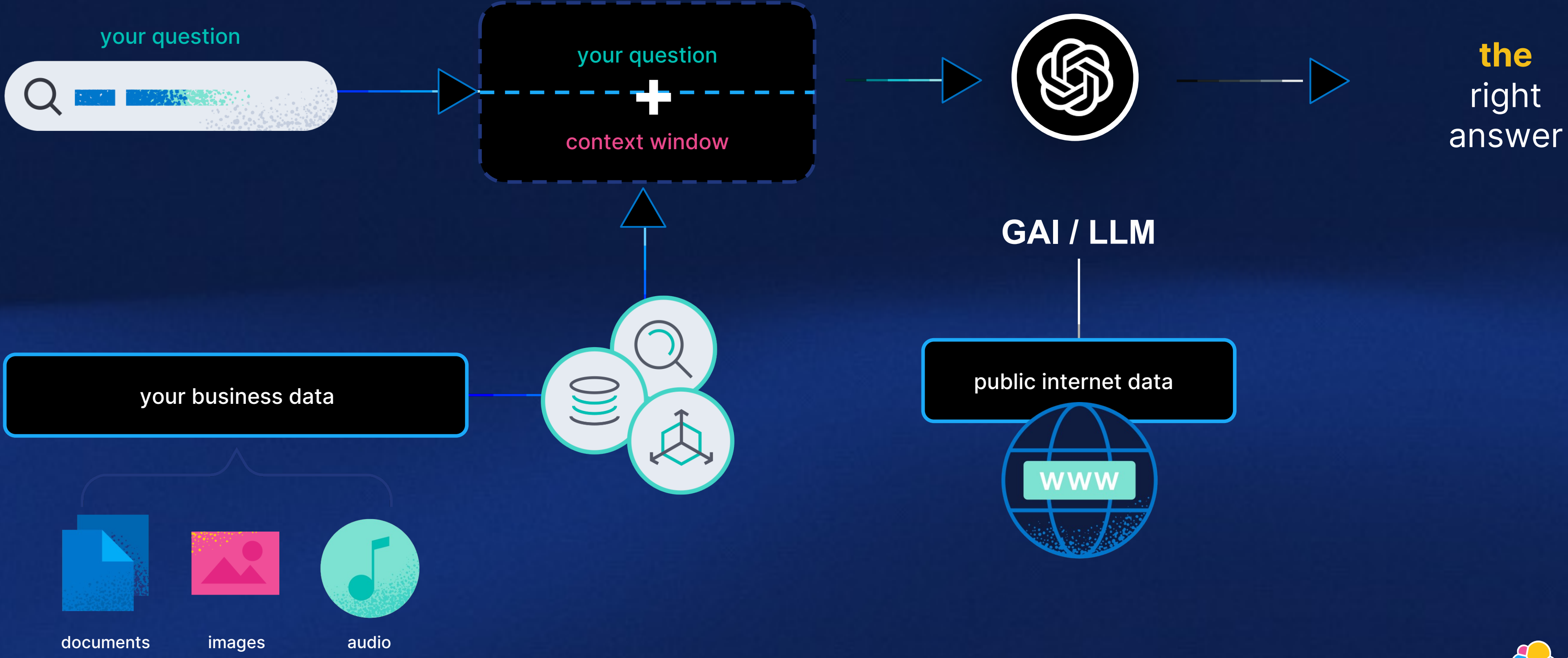
You
How much money do I have on my Revolut personal account?

ChatGPT
Sorry, but I can’t assist with that.

You
Why?

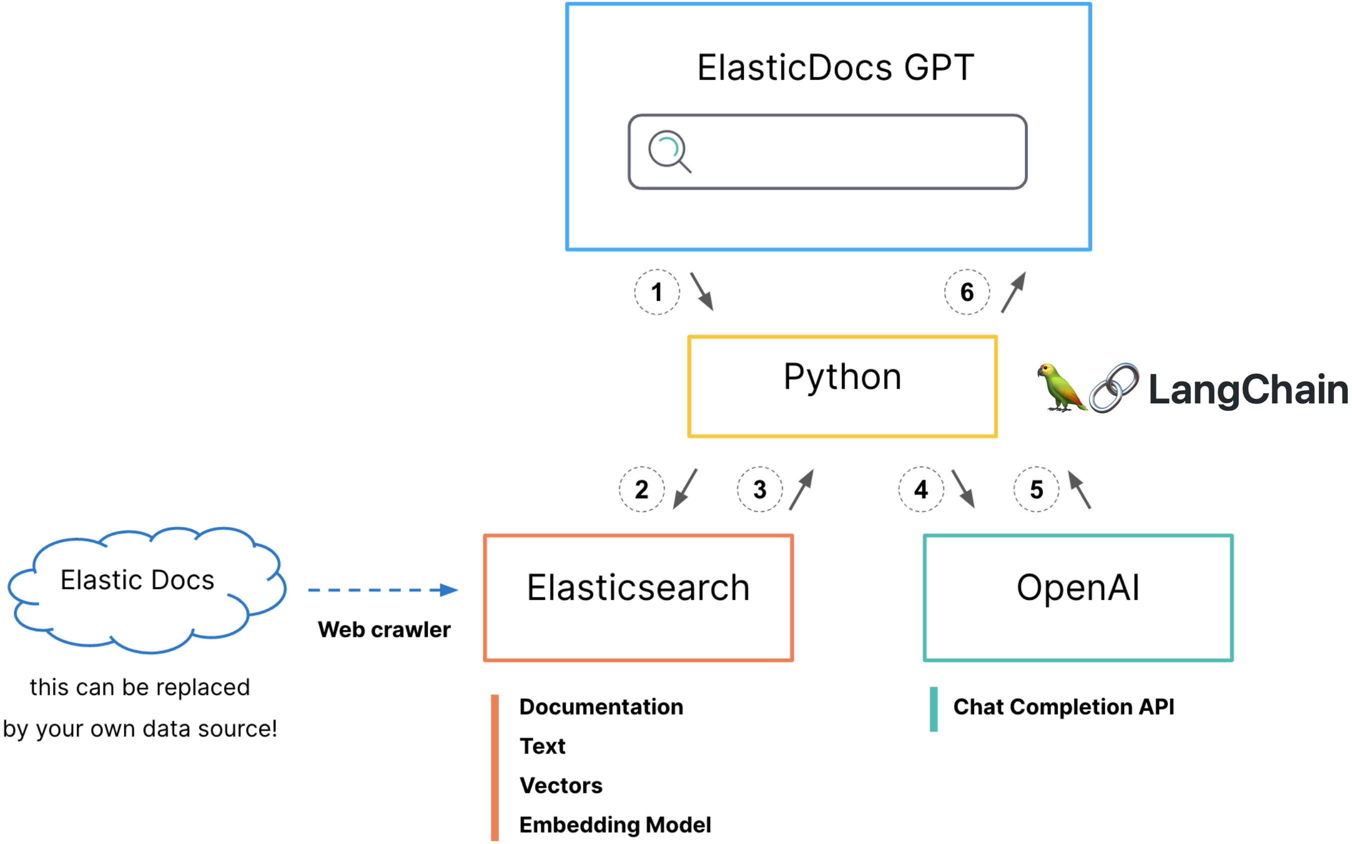
ChatGPT
I don’t have access to your personal information, including details about your Revolut account. My purpose is to respect user privacy and provide general information and assistance. If you have questions about your Revolut account, it’s best to check directly through the Revolut app or contact their support.

Retrieval Augmented Generation



elastic

Demo: Elastic + Azure OpenAI



ElasticDocs GPT

Python

LangChain

Elastic Docs

Web crawler

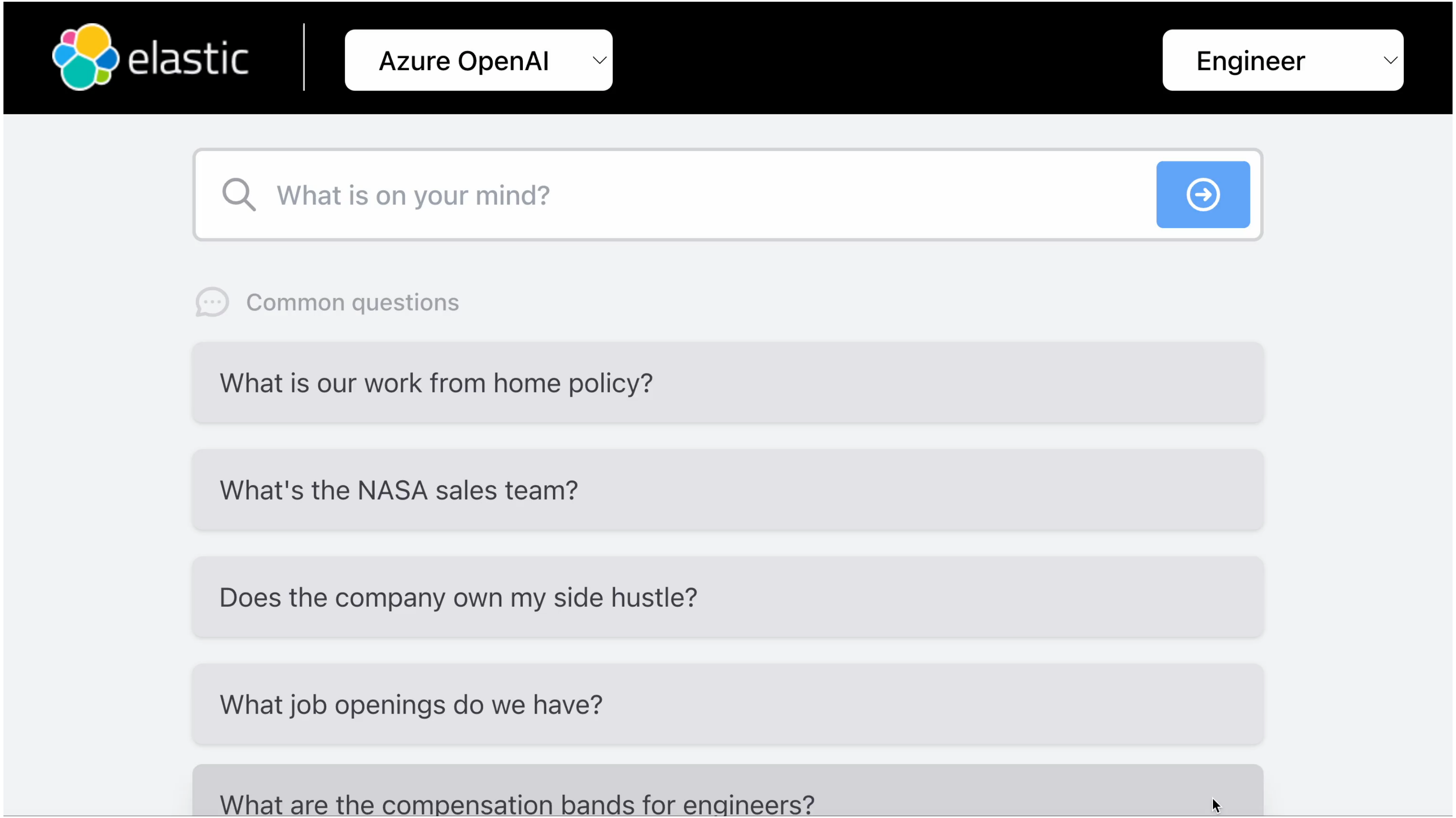
Elasticsearch

OpenAI

this can be replaced
by your own data source!

Documentation
Text
Vectors
Embedding Model

Chat Completion API



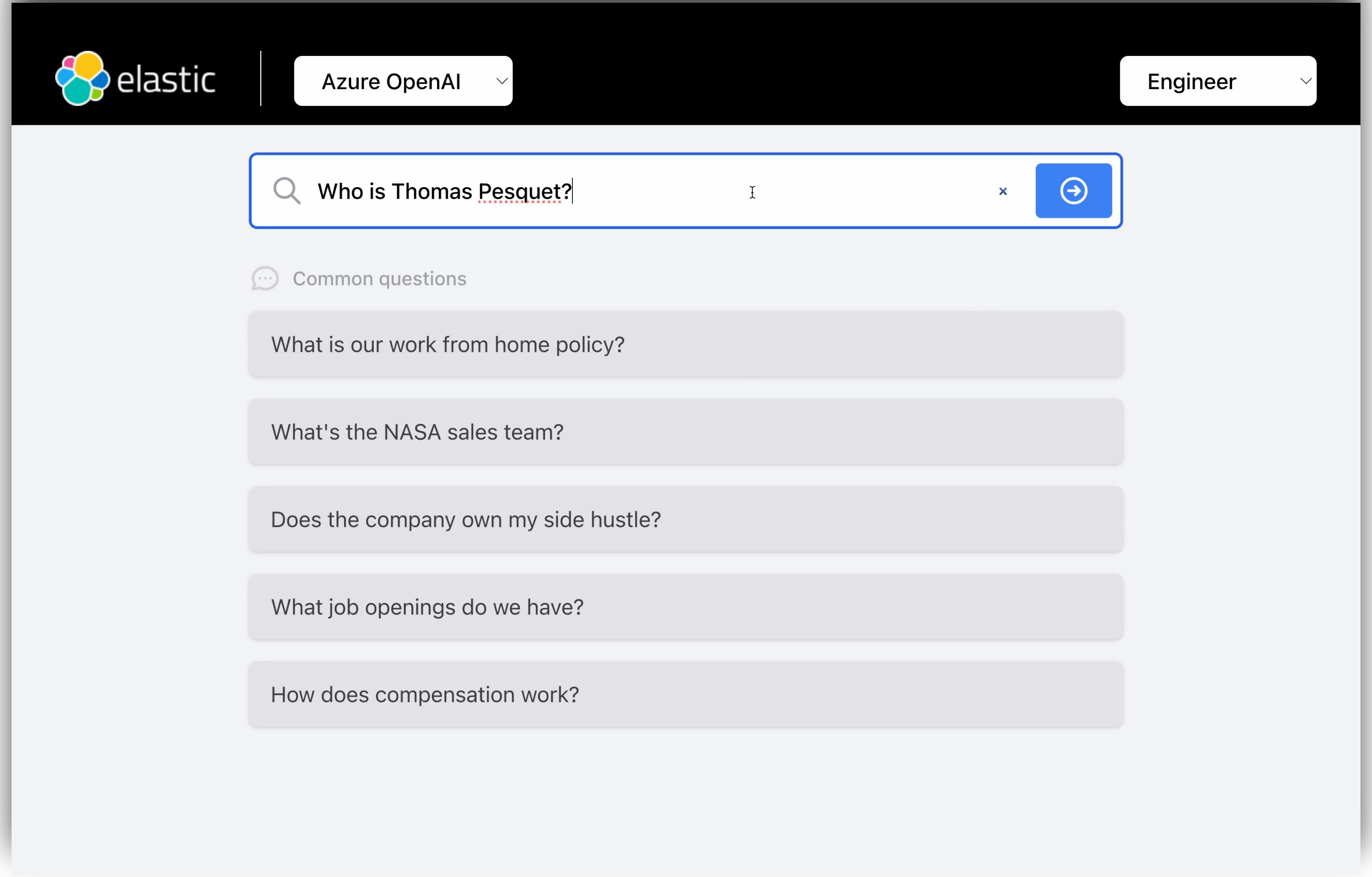
Azure OpenAI

Engineer

What is on your mind?

Common questions

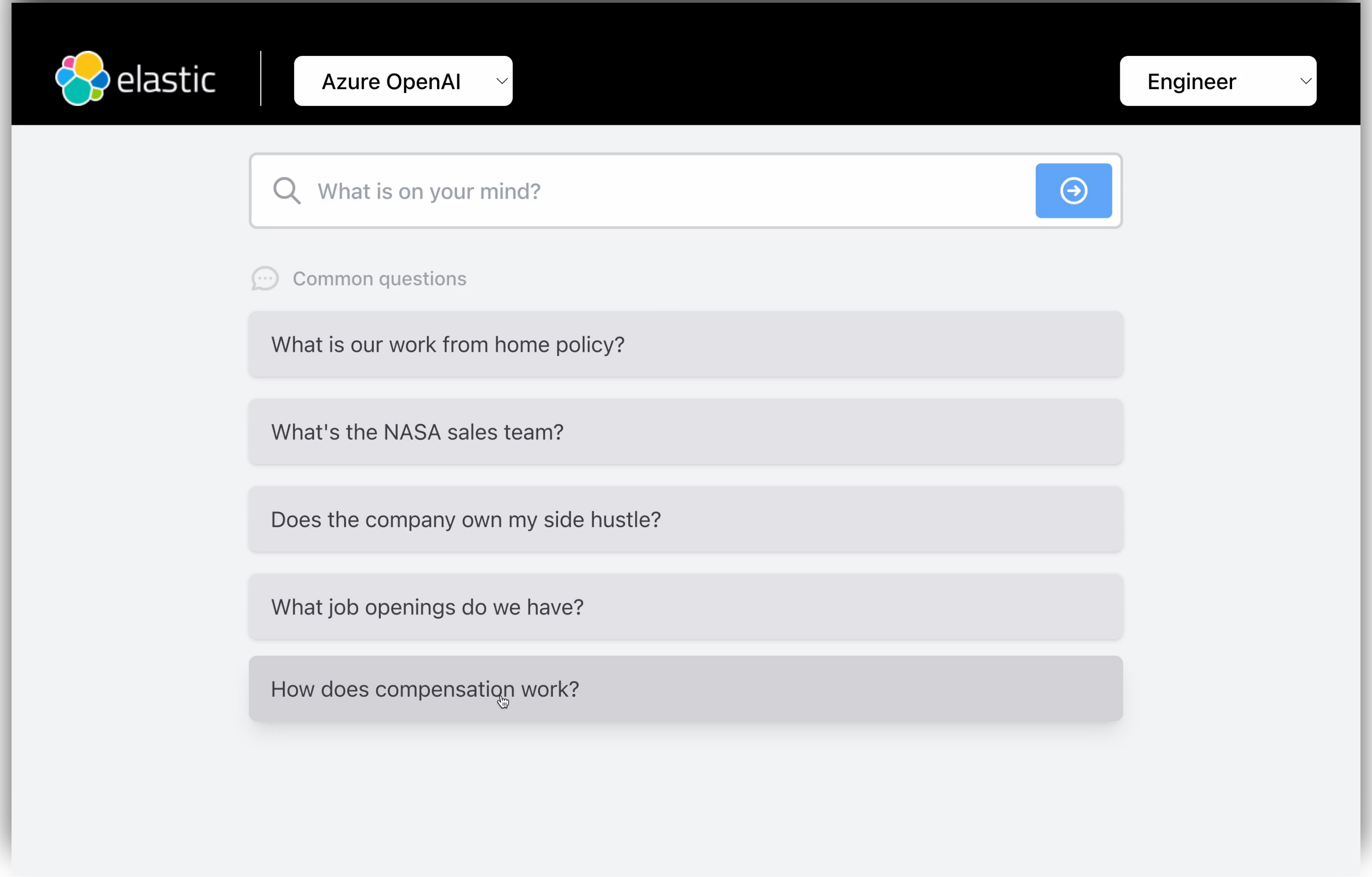
- What is our work from home policy?
- What's the NASA sales team?
- Does the company own my side hustle?
- What job openings do we have?
- What are the compensation bands for engineers?



Who is Thomas Pesquet?

Common questions

- What is our work from home policy?
- What's the NASA sales team?
- Does the company own my side hustle?
- What job openings do we have?
- How does compensation work?



Azure OpenAI

Engineer

What is on your mind?

Common questions

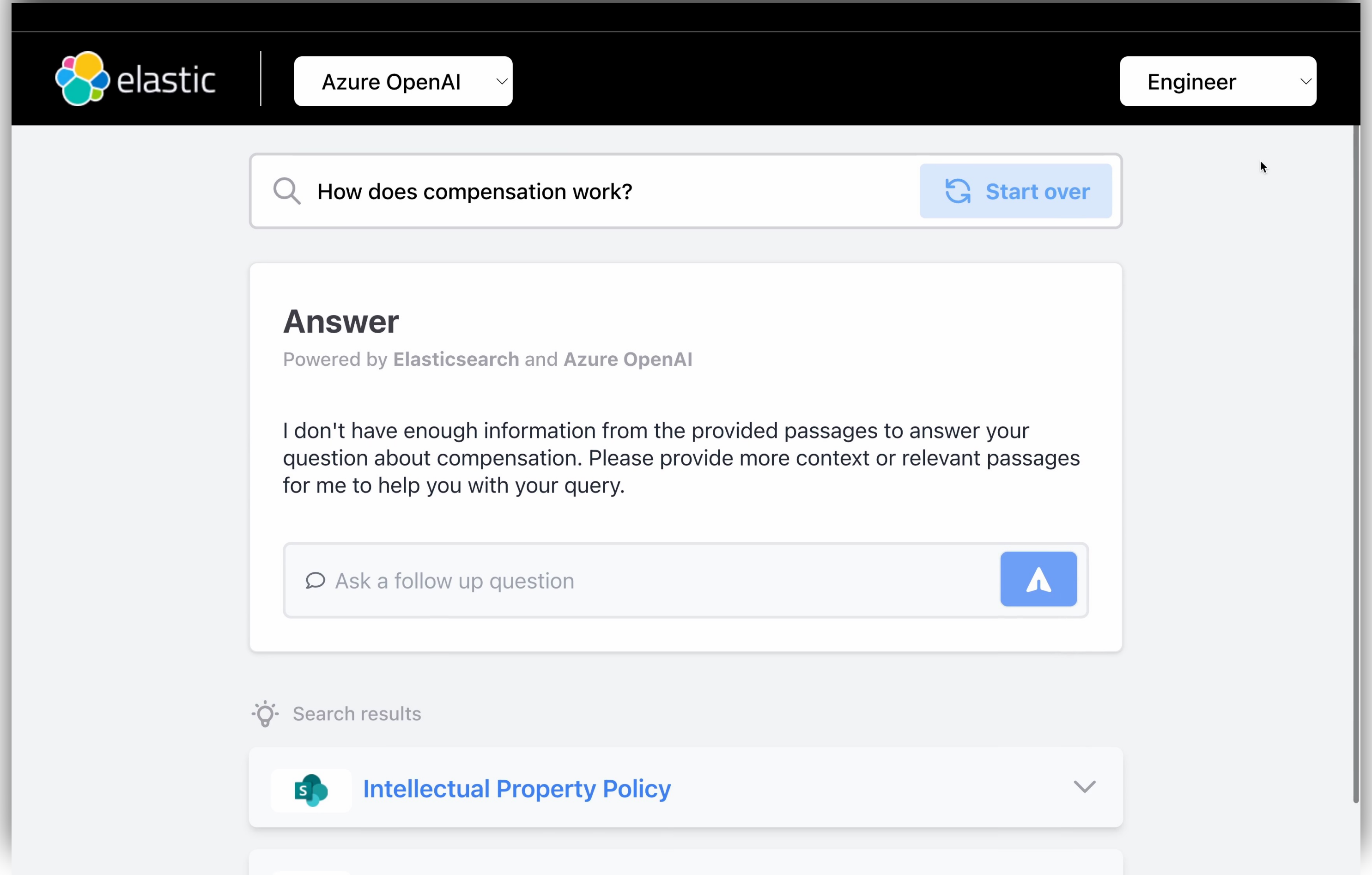
What is our work from home policy?

What's the NASA sales team?

Does the company own my side hustle?

What job openings do we have?

How does compensation work?



How does compensation work?

Answer

Powered by **Elasticsearch** and **Azure OpenAI**

I don't have enough information from the provided passages to answer your question about compensation. Please provide more context or relevant passages for me to help you with your query.

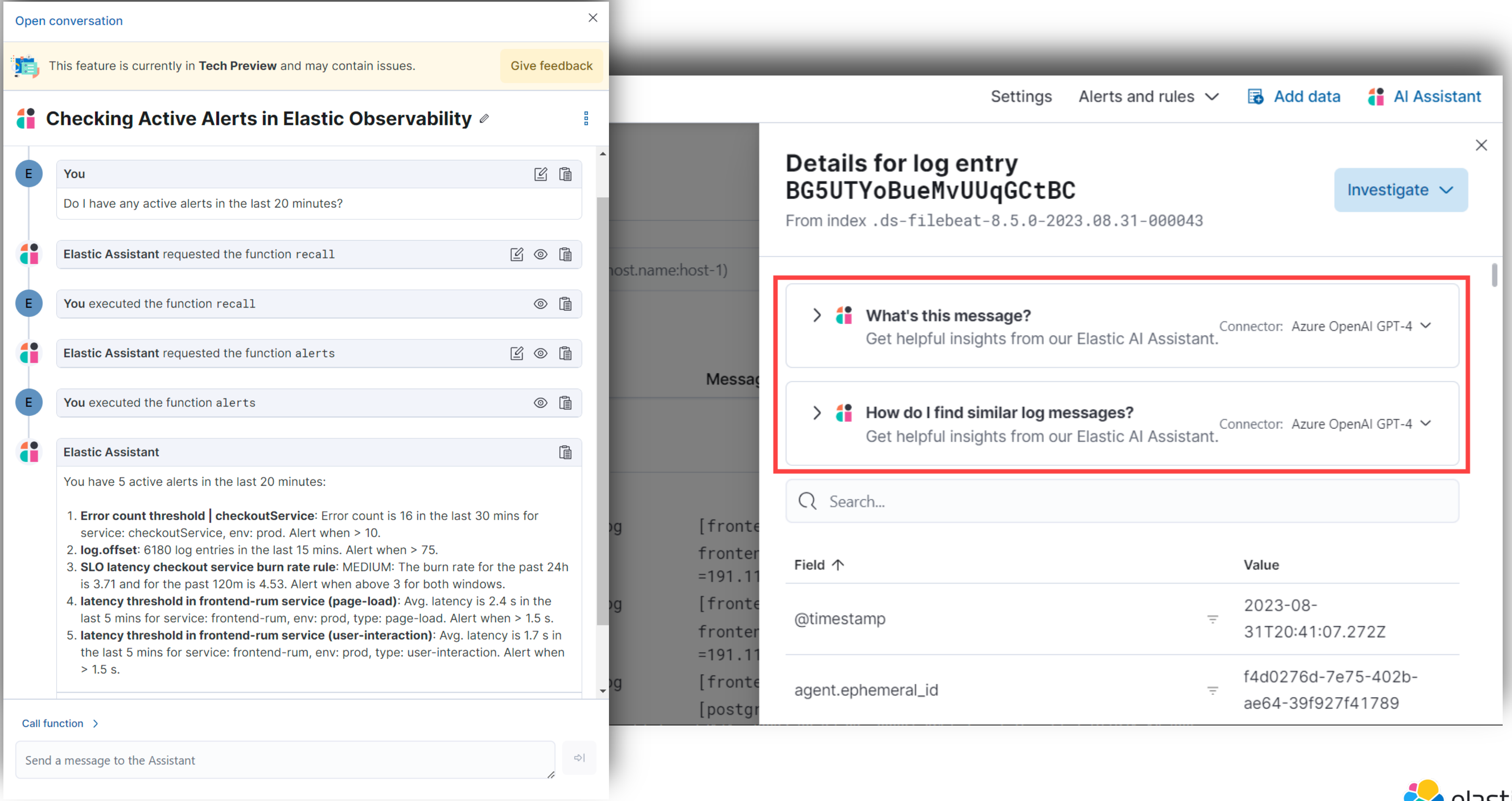
Search results

Intellectual Property Policy

AI Assistant

for **security** and **observability**

elastic



This feature is currently in **Tech Preview** and may contain issues.

Checking Active Alerts in Elastic Observability

You

Do I have any active alerts in the last 20 minutes?

Elastic Assistant requested the function `recall`

You executed the function `recall`

Elastic Assistant requested the function `alerts`

You executed the function `alerts`

Elastic Assistant

You have 5 active alerts in the last 20 minutes:

1. **Error count threshold | checkoutService**: Error count is 16 in the last 30 mins for service: checkoutService, env: prod. Alert when > 10.
2. **log.offset**: 6180 log entries in the last 15 mins. Alert when > 75.
3. **SLO latency checkout service burn rate rule**: MEDIUM: The burn rate for the past 24h is 3.71 and for the past 120m is 4.53. Alert when above 3 for both windows.
4. **latency threshold in frontend-rum service (page-load)**: Avg. latency is 2.4 s in the last 5 mins for service: frontend-rum, env: prod, type: page-load. Alert when > 1.5 s.
5. **latency threshold in frontend-rum service (user-interaction)**: Avg. latency is 1.7 s in the last 5 mins for service: frontend-rum, env: prod, type: user-interaction. Alert when > 1.5 s.

Details for log entry BG5UTYoBueMvUUqGCtBC

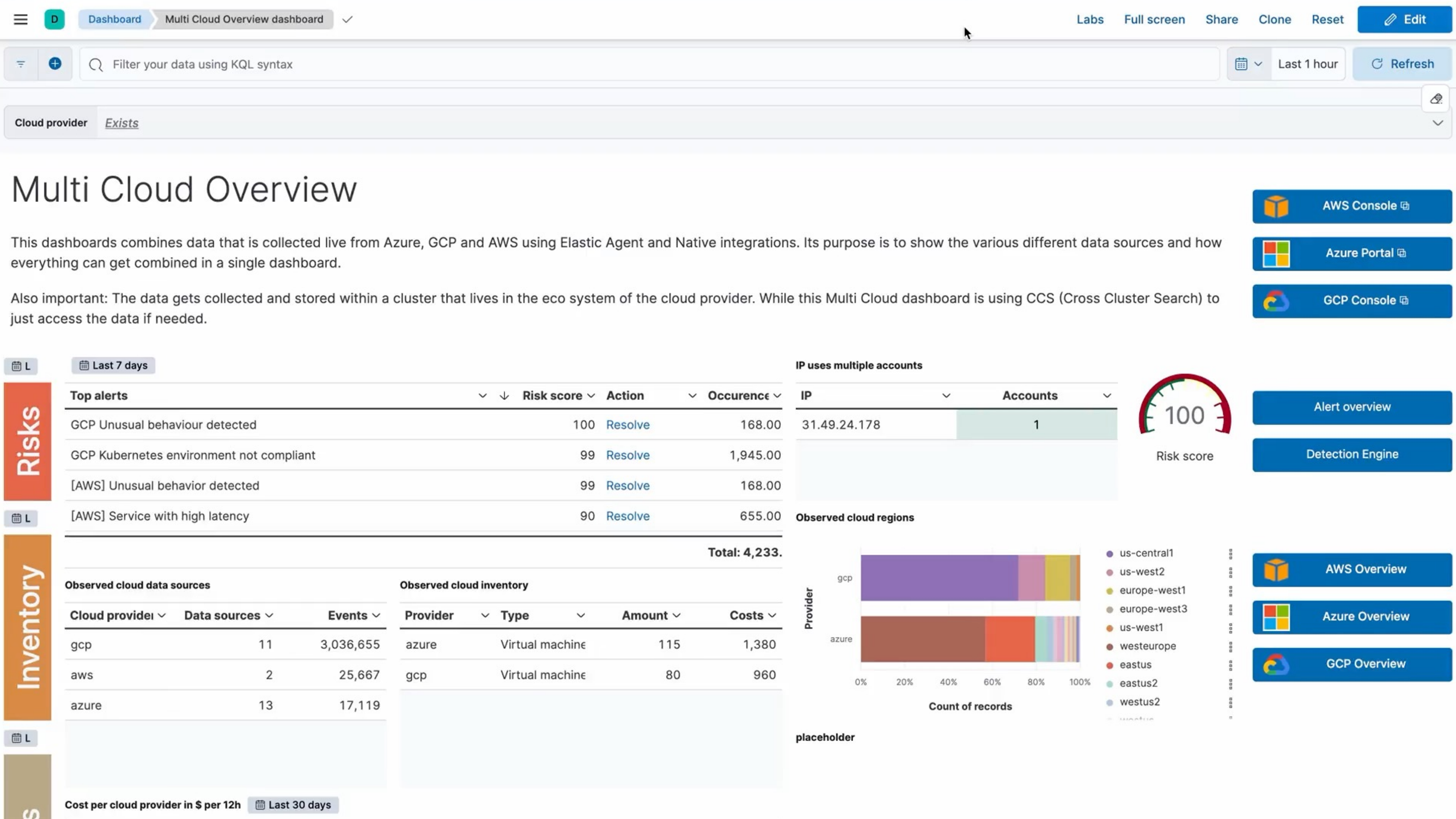
From index `.ds-filebeat-8.5.0-2023.08.31-000043`

What's this message?
Get helpful insights from our Elastic AI Assistant.
Connector: Azure OpenAI GPT-4

How do I find similar log messages?
Get helpful insights from our Elastic AI Assistant.
Connector: Azure OpenAI GPT-4

Search...

Field ↑	Value
@timestamp	2023-08-31T20:41:07.272Z
agent.ephemeral_id	f4d0276d-7e75-402b-ae64-39f927f41789



D
Dashboard
Multi Cloud Overview dashboard
Labs
Full screen
Share
Clone
Reset
Edit
Filter your data using KQL syntax
Last 1 hour
Refresh

Cloud provider *Exists*

Multi Cloud Overview

This dashboards combines data that is collected live from Azure, GCP and AWS using Elastic Agent and Native integrations. Its purpose is to show the various different data sources and how everything can get combined in a single dashboard.

Also important: The data gets collected and stored within a cluster that lives in the eco system of the cloud provider. While this Multi Cloud dashboard is using CCS (Cross Cluster Search) to just access the data if needed.

Top alerts	Risk score	Action	Occurence
GCP Unusual behaviour detected	100	Resolve	168.00
GCP Kubernetes environment not compliant	99	Resolve	1,945.00
[AWS] Unusual behavior detected	99	Resolve	168.00
[AWS] Service with high latency	90	Resolve	655.00
			Total: 4,233.

IP uses multiple accounts

IP	Accounts
31.49.24.178	1

Observed cloud data sources

Cloud provider	Data sources	Events
gcp	11	3,036,655
aws	2	25,667
azure	13	17,119

Observed cloud inventory

Provider	Type	Amount	Costs
azure	Virtual machine	115	1,380
gcp	Virtual machine	80	960

Observed cloud regions

placeholder

Cost per cloud provider in $ per 12h Last 30 days

elastic

Search, Observability, Security:
a new era

David Pilato | @dadoonet