

# Neuronale Suche mit Elasticsearch

- 01 Einführung & Terminologie
- 02 Wie funktioniert eine Text-Suche?
- 03 Neuronale Suche
- 04 Fazit

# *Einführung*

Einführung in Konzepte  
und Terminologie

Keine Demo

Disclaimer

KI-Suche / ChatGPT  
als «Hot Topic»

Neue Produkte. Überall!

Elasticsearch im  
Technologie-Portfolio

«Elasticsearch ist  
nicht mehr relevant»

Wieso?

# Terminologie

**Term:** Was Benutzer:innen in eine Suchfeld eingeben.

**Dokument:** Inhalte die durchsucht werden sollen (Webseite, News, ...)

**Index:** Kollektion von Dokumenten

# Terminologie: Suche

**Vektor:** Format um Daten  
abzuspeichern.

**KNN:** Algorithmus um Daten  
basierend auf Ähnlichkeit zu  
finden.

# Terminologie: Vektor

### **Language Model (LM):**

Language Model try to determine the probability of a sequence of words using statistics and probability techniques

### **Large Language Model (LLM):**

Language Model consisting of Neuronal Networks with Billions of weights.

Examples: Bert (Google), GPT (OpenAI)

# Terminologie: Model

TF / IDF – Wie funktioniert eine Text-Suche?

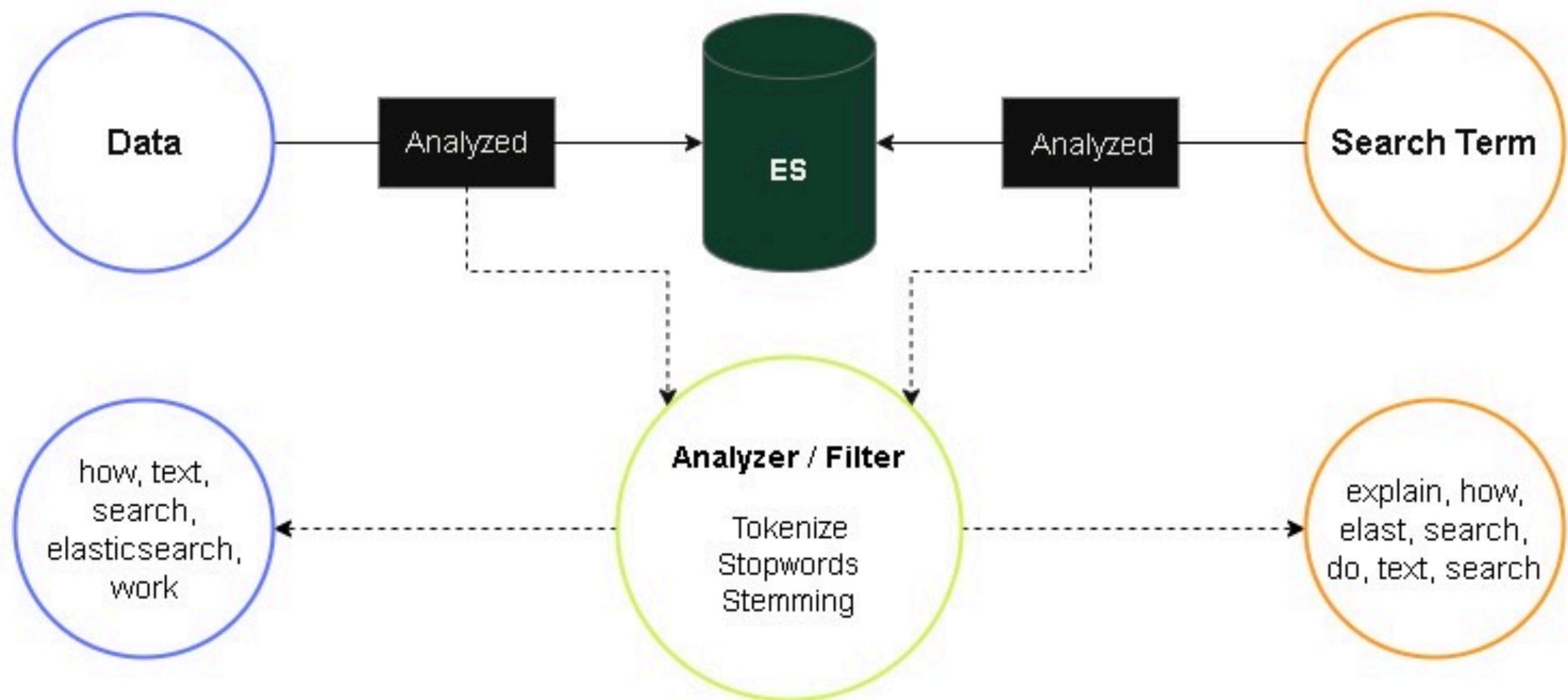
TF / IDF – Wie relevant ist ein Dokument?

**Term Frequency** – Je häufiger ein Term in einem Dokument vorkommt, desto relevanter ist ein Dokument.

**Inverse Document Frequency** – Je weniger andere Dokumente den Term enthalten desto relevanter ist ein Dokument.

## How Text Search in Elasticsearch works

## Explain me how Elastic Search is doing Text Search



# Diskussion

+ Schnell und geringe(re) HW-Kosten

+ Transparent

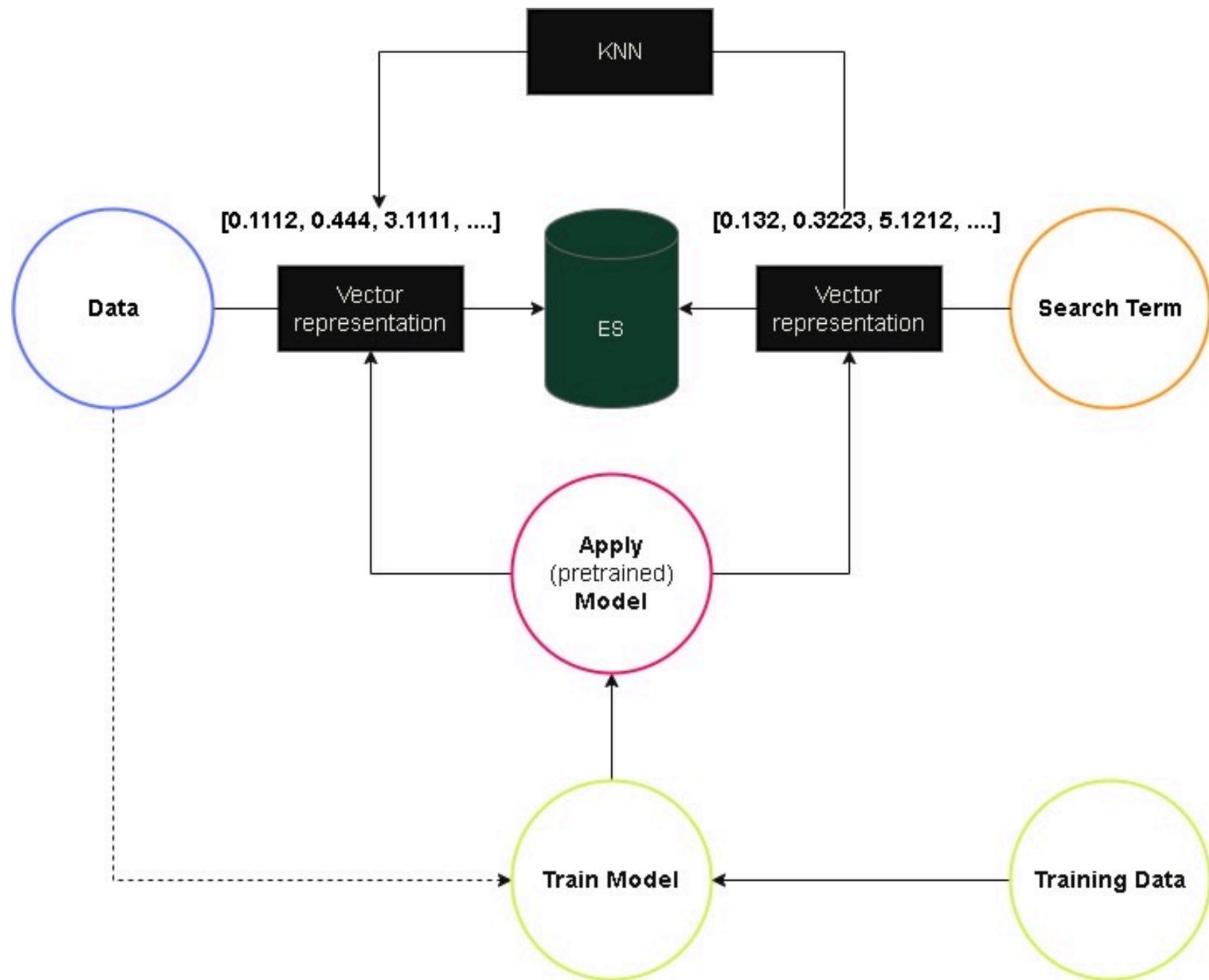
+ Oft gut genug

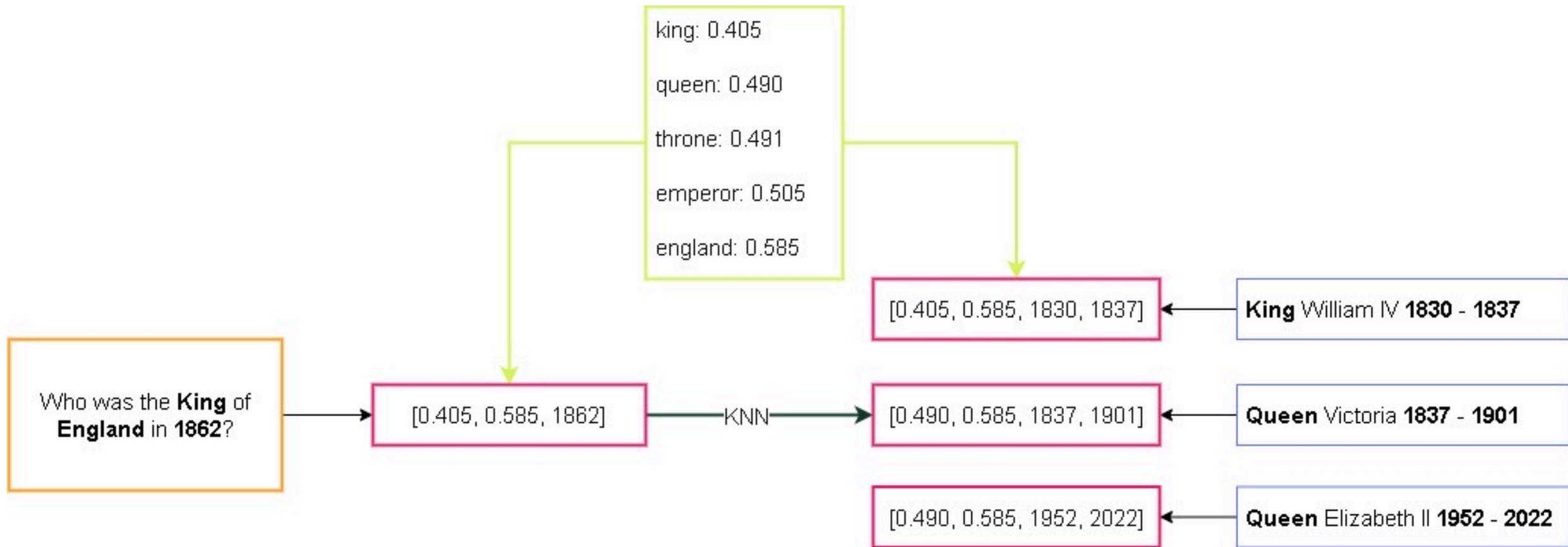
- Nur für Text

- In gewissen Anwendungsfällen limitiert

elast <-> elasticsearch

# Neuronale Suche





# Diskussion

+ Smart(er)

+ Erweiterte Anwendungsfälle

PDF, Bilder, Kontext

- Hohe Anforderungen an Hardware

- Komplexität der Modelle

- Abhängigkeit von (guten) Datensets

# Hybrid-Suche

TF/IDF und neuronale Suche können kombiniert werden!

Beschaffung von Trainingsdaten  
für spezifische Anwendungsfälle  
ist aufwändig.

**Aufwand / Ertrag** mit  
TF/IDF potenziell besser für  
fachspezifisches Vokabular.

# Hybrid-Suche

(Zwischen)Fazit

# 01

Chatbot

Chatbot-Suche mit Elasticsearch  
ist möglich

Elasticsearch ist «Träger». Es wird  
keine Integrierte LLMs / KI in  
Elasticsearch geben.

# 02

Verfügbarkeit der Modelle

Modelle sind wichtig.

OpenSource Modelle sind vorhanden;  
Sind meist generisch; Trainingsdaten für  
Deutsch rar.

# 03

Komplexität und Hardware

Komplexität und Anforderungen an  
die Hardware steigen

Herausforderungen: Resultate erklären,  
Modelle trainieren, Quellen-Angaben,  
Performance

Vielen Dank für Ihre Aufmerksamkeit.