Vector Search 101

UNDERSTANDING EMBEDDINGS, SIMILARITY, AND SEARCH AT SCALE

PRESENTED BY: GIRISH NAUTIYAL

SENIOR DATA SCIENTIST

What is Vector Search?

Traditional Search

cat

catalog

category

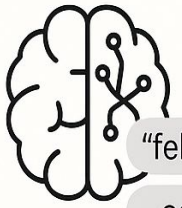
Matches keywords or exact terms

Query: “cheap laptop’ → Only exact word match

Rule-based, SQL, filters

Fails on synonyms or rephrasing

Vector Search



“feline”

cat

kitten

pet

Uses semantic embeddings to match meaning

Query: “budget notebook” → Returns "cheap laptop”

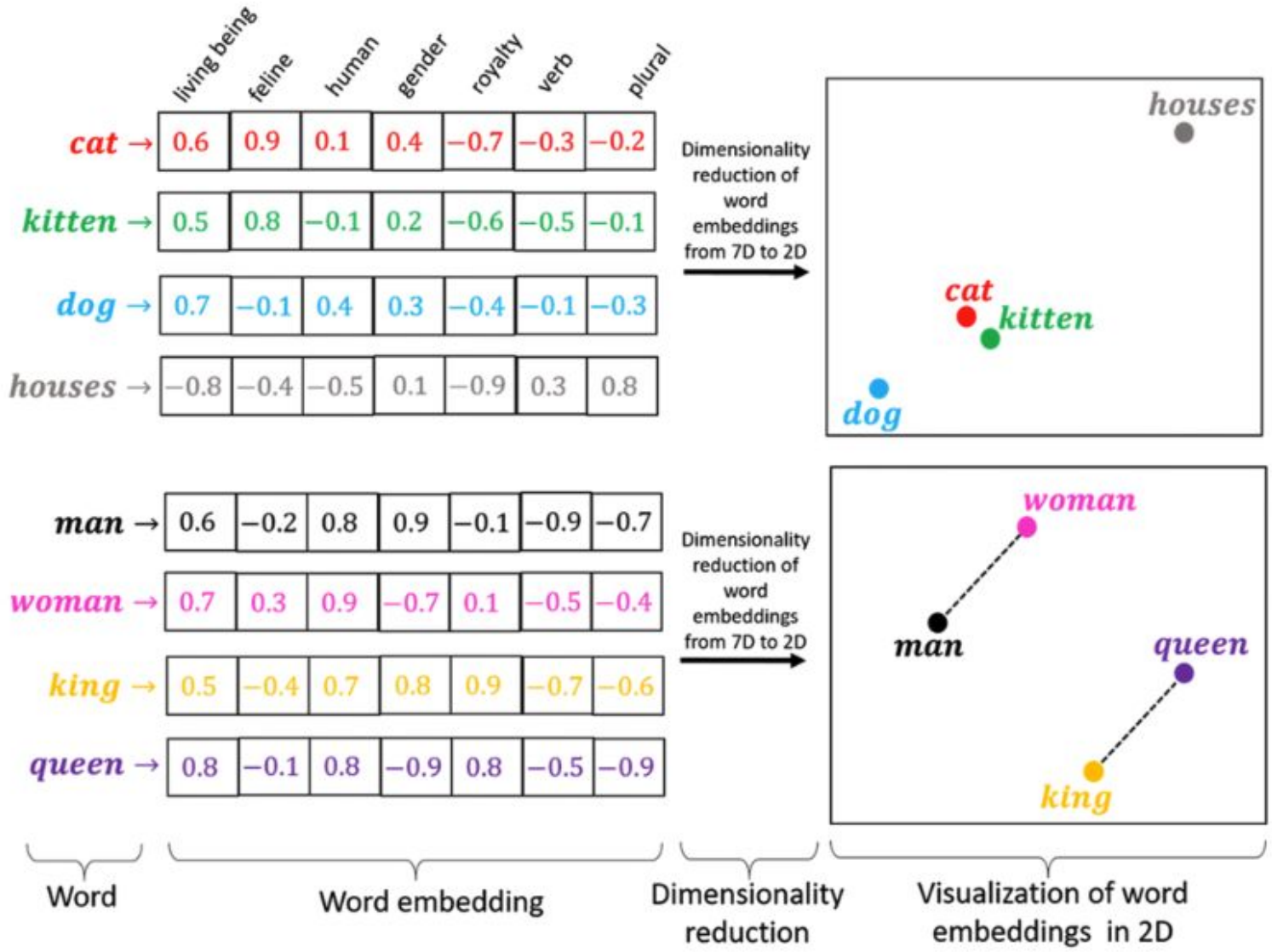
Vector similarity (e.g. cosine)

Requires model + compute resources

Real-world Use Cases

- Semantic search
  ("quiet places to work" → libraries, cafes)
- Product recommendations
- Image & video search
- Question answering (RAG systems)

What are Embeddings?

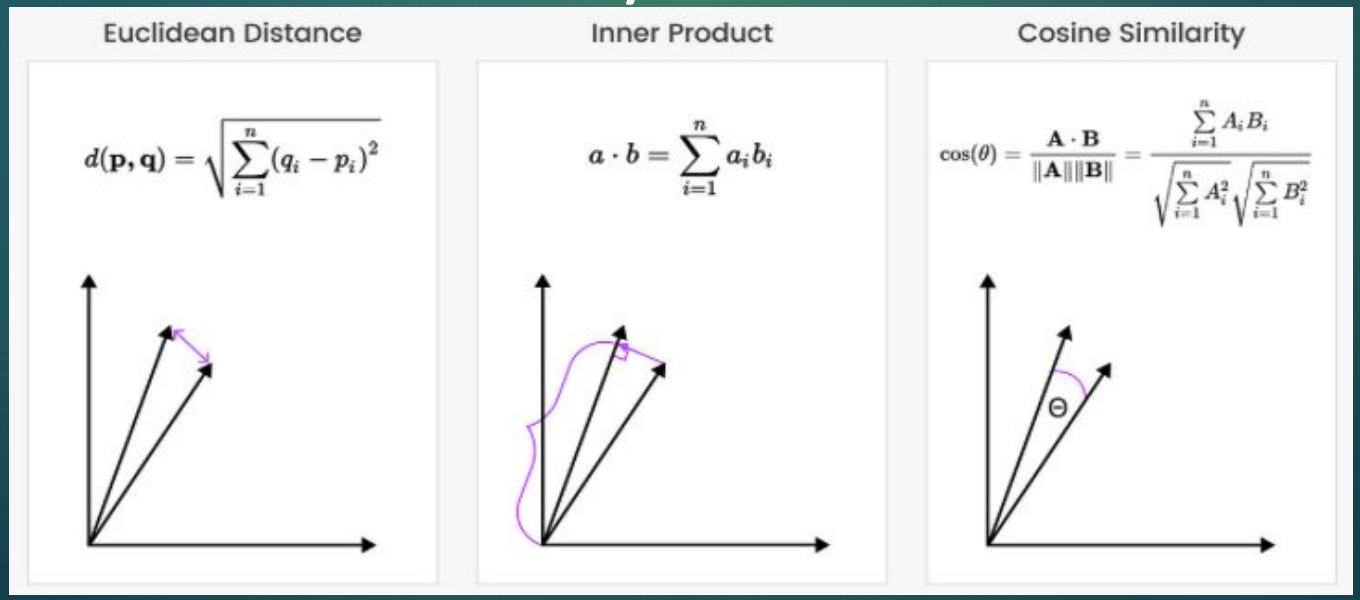


Embedding Generation Models

- Popular models:
  - BERT variants
  - OpenAI Embeddings
  - Sentence Transformers library
- Use APIs or Hugging Face models

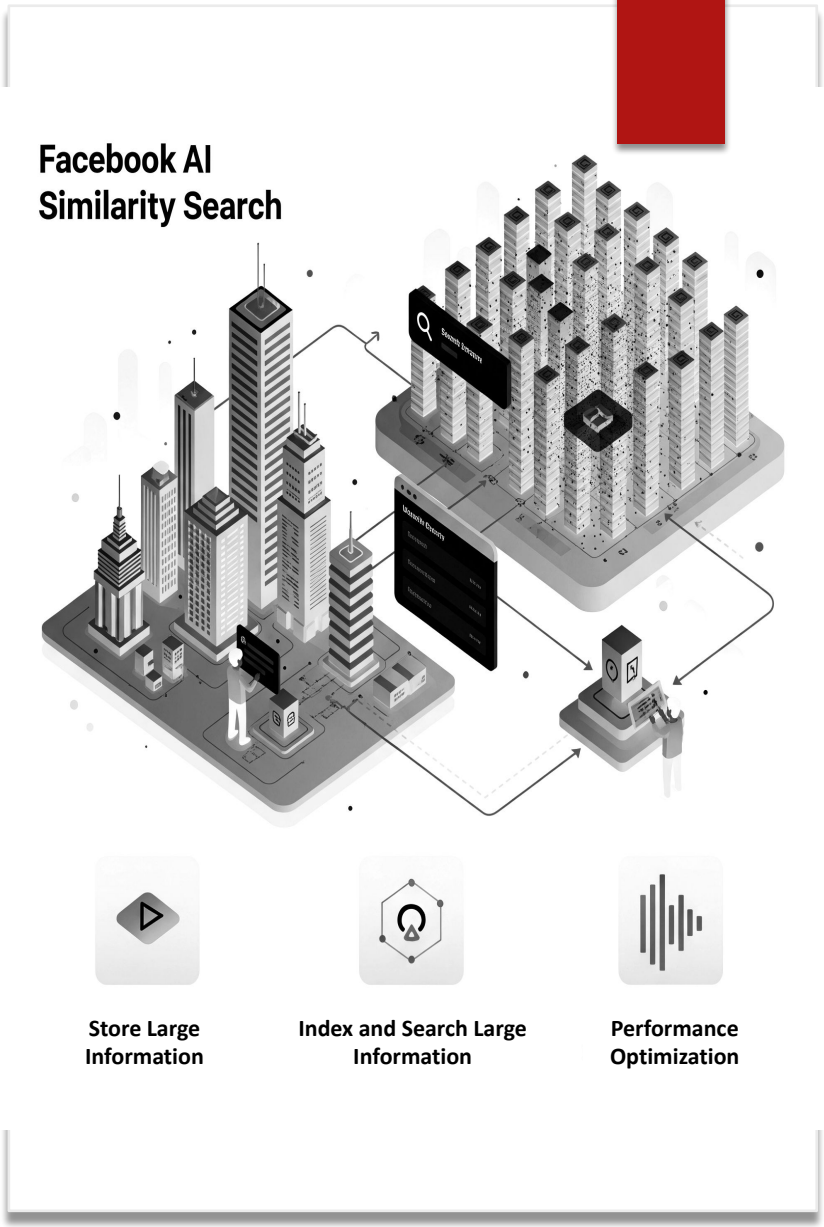
Similarity Metrics

- Measure how close two vectors are
- Common metrics:
  - Euclidean Distance
  - Dot Product
  - Cosine Similarity



What is FAISS?

- Facebook AI Similarity Search
- Library to store, index, and search large vector collections
- Optimized for performance (millions of vectors)



Putting it All Together

Step-by-step pipeline:

- Convert data to embeddings
- Similarity Matrix
- Store in FAISS index
- Search new query vector
- Retrieve and show results

Let's Code

Embeddings	Embed text using Sentence Transformers
Similarity	Understand Similarity
FAISS	Work with FAISS
Results	Analyze results

Final Thoughts

- Vector search unlocks semantic understanding
- Useful in modern AI systems
- Learn and experiment using Hugging Face + FAISS

Q&A

- Any questions?
- Feel free to ask about:
  - Use cases
  - Tools
  - Getting started projects