

Embeddings

Computing

Turning text into vectors so that meaning becomes distance.

A model that maps a piece of text to a list of numbers — a vector — positioned so that texts with similar meanings sit close together in that space.

COMMON SIZE
768–1536

METRIC
cosine

RANGE
–1 to 1

- Typical sizes run from 384 to 3,072 dimensions.
- Similarity is usually cosine distance between vectors.
- Two paraphrases land close together even with no shared words.
- Used for search, clustering, deduplication and recommendation.
- The same model must embed both query and documents.
- Changing embedding model means re-embedding everything.

LOOK FOR IT

'How do I reset my password' matches a document titled 'Account recovery' with no word in common.

Embeddings capture similarity, not truth. Two confidently opposite statements can sit very close together.