

Vector Databases and Similarity Search

What is a Vector Database?

A vector database stores embeddings for fast semantic search.

Embeddings

Numerical representations of meaning produced by ML models.

Similarity Metrics

Cosine similarity, dot product, and Euclidean distance are commonly used.

Indexing Techniques

Approximate Nearest Neighbor (ANN) indexes improve retrieval speed.

Role in RAG

Vector databases retrieve context documents for LLM generation.