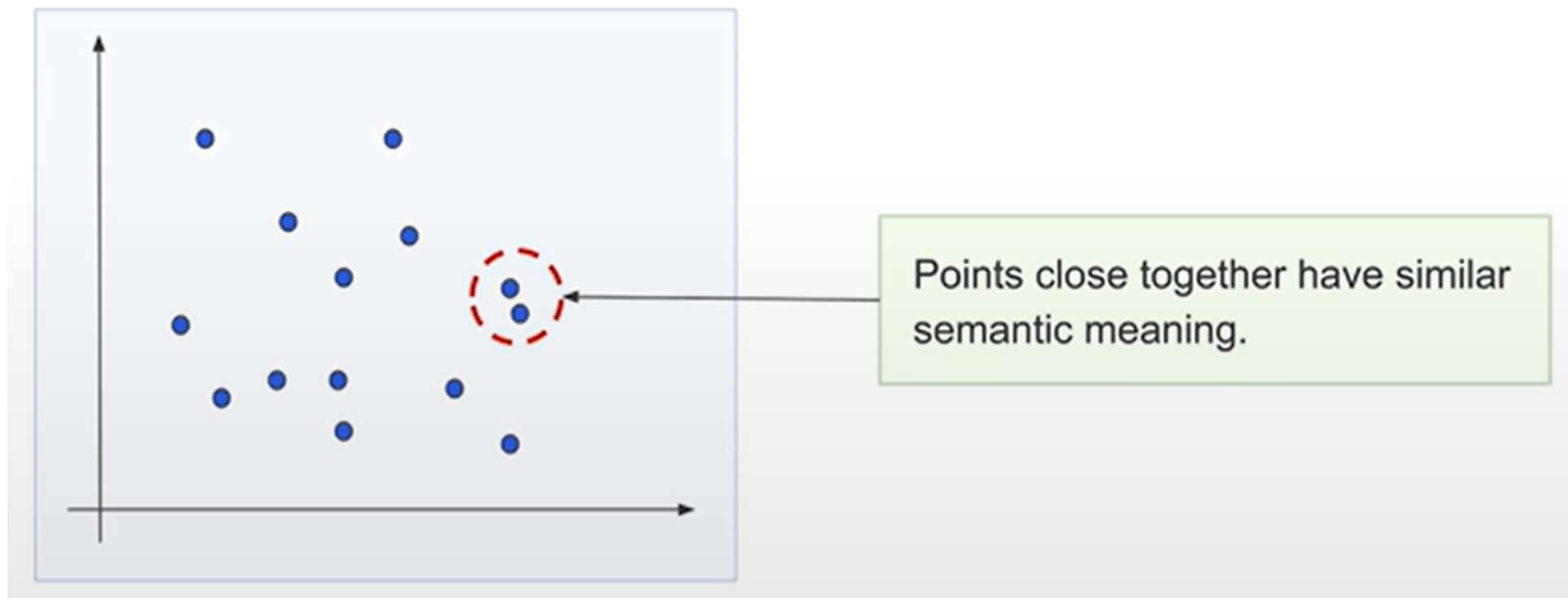

SENTENCE EMBEDDING AND APPLICATIONS

Minha Hwang



What are Vector Embedding?



Source: Deeplearning.ai – Embedding Models from Architecture to Implementation

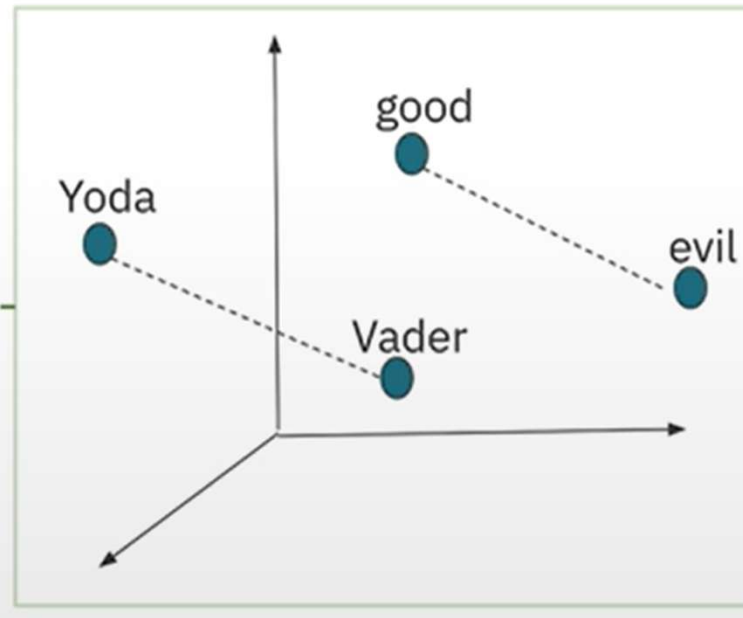
<https://learn.deeplearning.ai/courses/embedding-models-from-architecture-to-implementation>

Word Embeddings

Word-embeddings: Word2vec, Glove, others

Word2vec Algebra

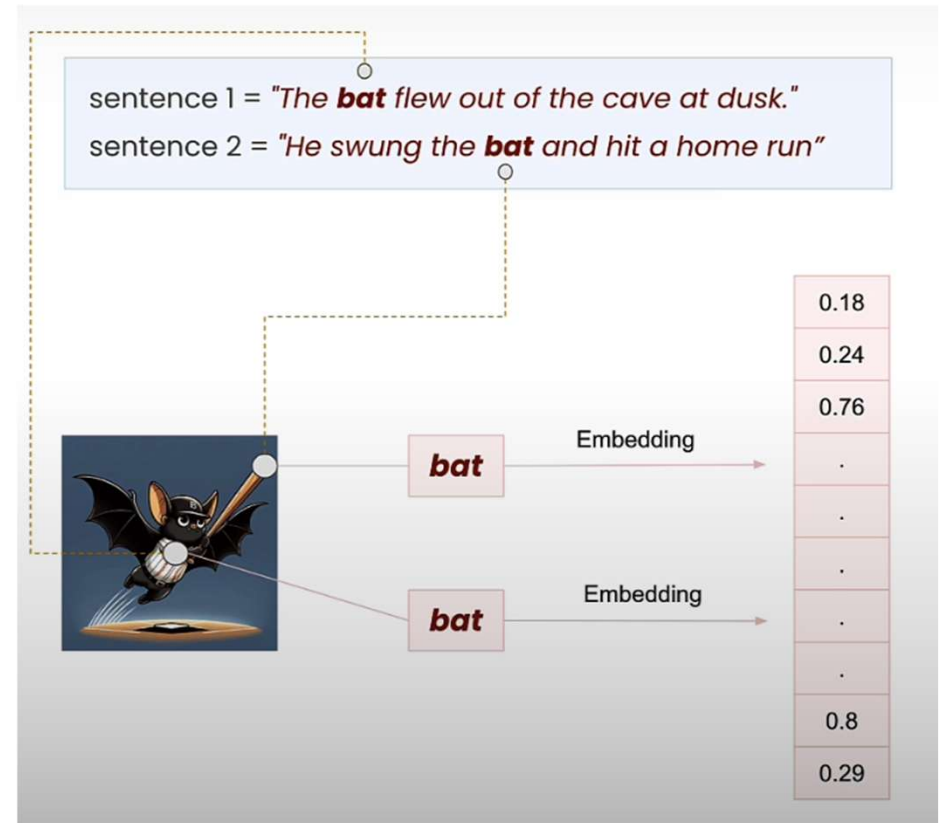
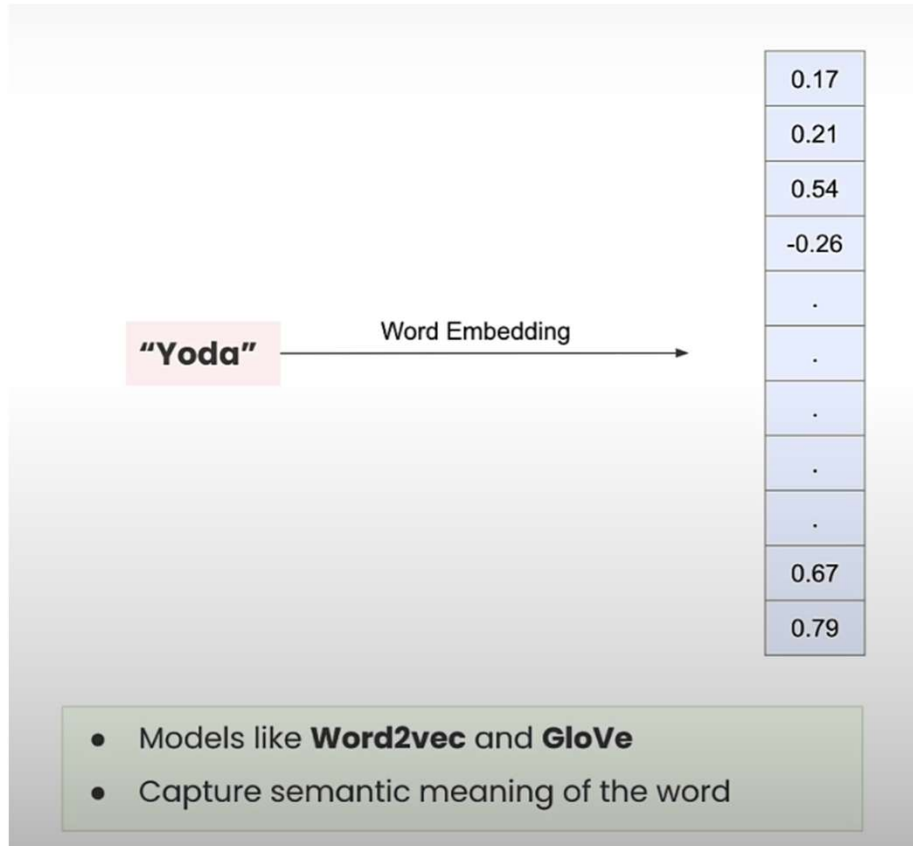
$\text{emb}(\text{"Vader"}) = \text{emb}(\text{"Yoda"}) - \text{emb}(\text{"good"}) + \text{emb}(\text{"evil"})$



Source: Deeplearning.ai – Embedding Models from Architecture to Implementation

<https://learn.deeplearning.ai/courses/embedding-models-from-architecture-to-implementation>

Problems with Word Embedding: Context are Lost!



Source: Deeplearning.ai – Embedding Models from Architecture to Implementation

<https://learn.deeplearning.ai/courses/embedding-models-from-architecture-to-implementation>

Sentence Embedding: Captures Semantic Meaning

Sentence Embeddings: Dense vector representations of sentences that capture their semantic meaning.

"I have a dream that one day this nation will rise up and live out the true meaning of its creed: We hold these truths to be self-evident, that all men are created equal."

Sentence



Embedding vector

Source: Deeplearning.ai – Embedding Models from Architecture to Implementation

<https://learn.deeplearning.ai/courses/embedding-models-from-architecture-to-implementation>

Embedding Vectors in Other Domains

Image embeddings translate visual content into a vector form.

Video embeddings capture the characteristics of video data, including visual appearance and temporal dynamics.

Audio embeddings represent sound signals in a vector space.

Graph embeddings convert nodes, edges, and their features in a graph into vector space, preserving structural information.

Shared embedding spaces: For example, **CLIP** which generates embeddings for both text and images

Source: Deeplearning.ai – Embedding Models from Architecture to Implementation

<https://learn.deeplearning.ai/courses/embedding-models-from-architecture-to-implementation>

Applications of Embeddings

LLMs: Input tokens are converted to token embeddings.

Semantic Search: Enhances search engines by retrieving sentences with similar meanings, improving search relevance.

RAG: Sentence embeddings enable efficient retrieval of relevant chunks

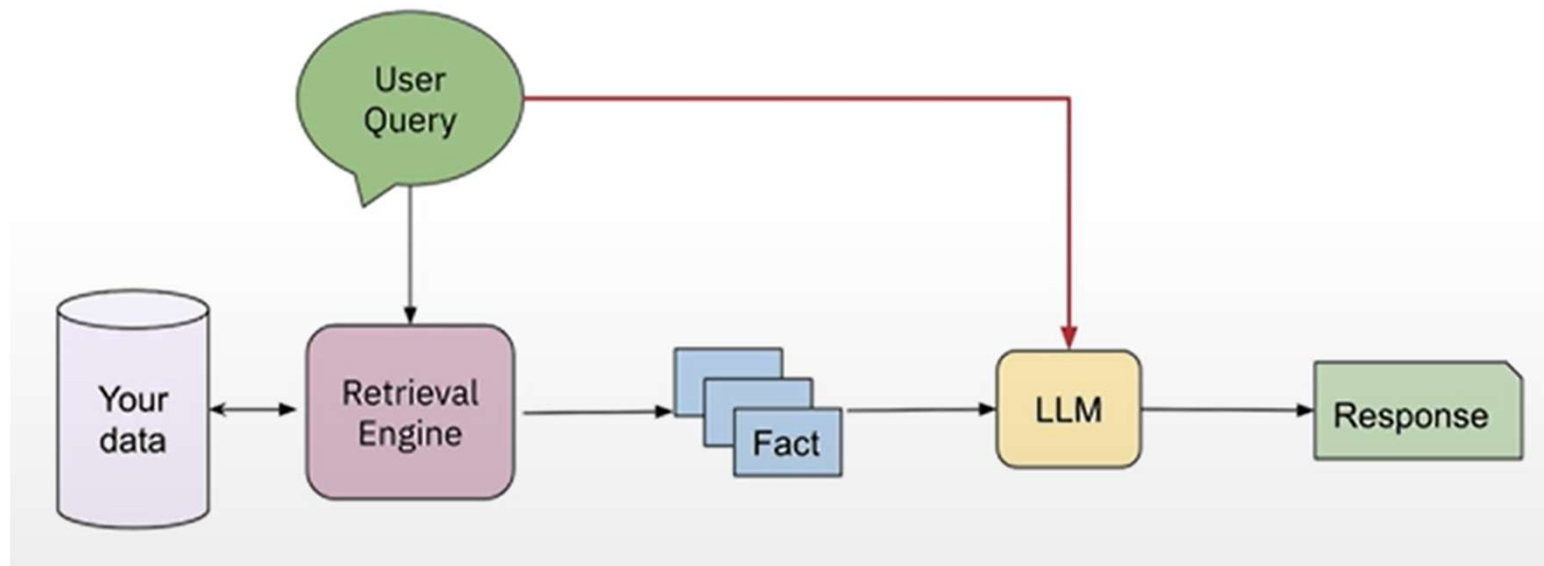
Recommendations: Representing products in embedding space and using similarity search.

Anomaly Detection: Identifies patterns in data that deviate significantly from the norm.

Source: Deeplearning.ai – Embedding Models from Architecture to Implementation

<https://learn.deeplearning.ai/courses/embedding-models-from-architecture-to-implementation>

Sentence Embedding: Critical for RAG



Source: Deeplearning.ai – Embedding Models from Architecture to Implementation

<https://learn.deeplearning.ai/courses/embedding-models-from-architecture-to-implementation>

Sentence Embedding Models: SBERT, Short Sentence Transformer

USE (Universal Sentence Encoder) by Google

SBERT (Sentence Bert) – Nils Rimers and Iryna Gurevych

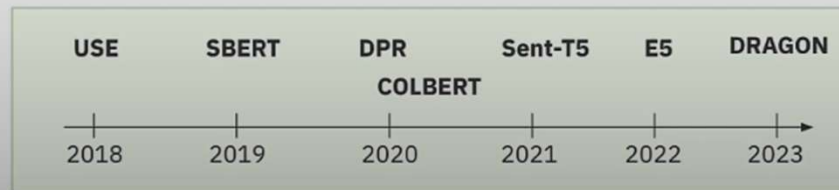
DPR (dense passage retrieval) by facebook and others

Sentence-T5: a variant of the famous T5 model specifically trained for sentence embedding

E5: Using signals from curated multi-domain datasets to achieve SOTA sentence embedding model performance.

DRAGON: A technique that uses large scale synthetic data and curriculum learning to improve retrieval performance with sentence embedding models

COLBERT: Using multiple vectors to represent sentence embeddings and “late” interaction



Source: Deeplearning.ai – Embedding Models from Architecture to Implementation

<https://learn.deeplearning.ai/courses/embedding-models-from-architecture-to-implementation>

Demo

- **Semantic Query Reformulation**
- **Topic Modeling: Query Clustering**