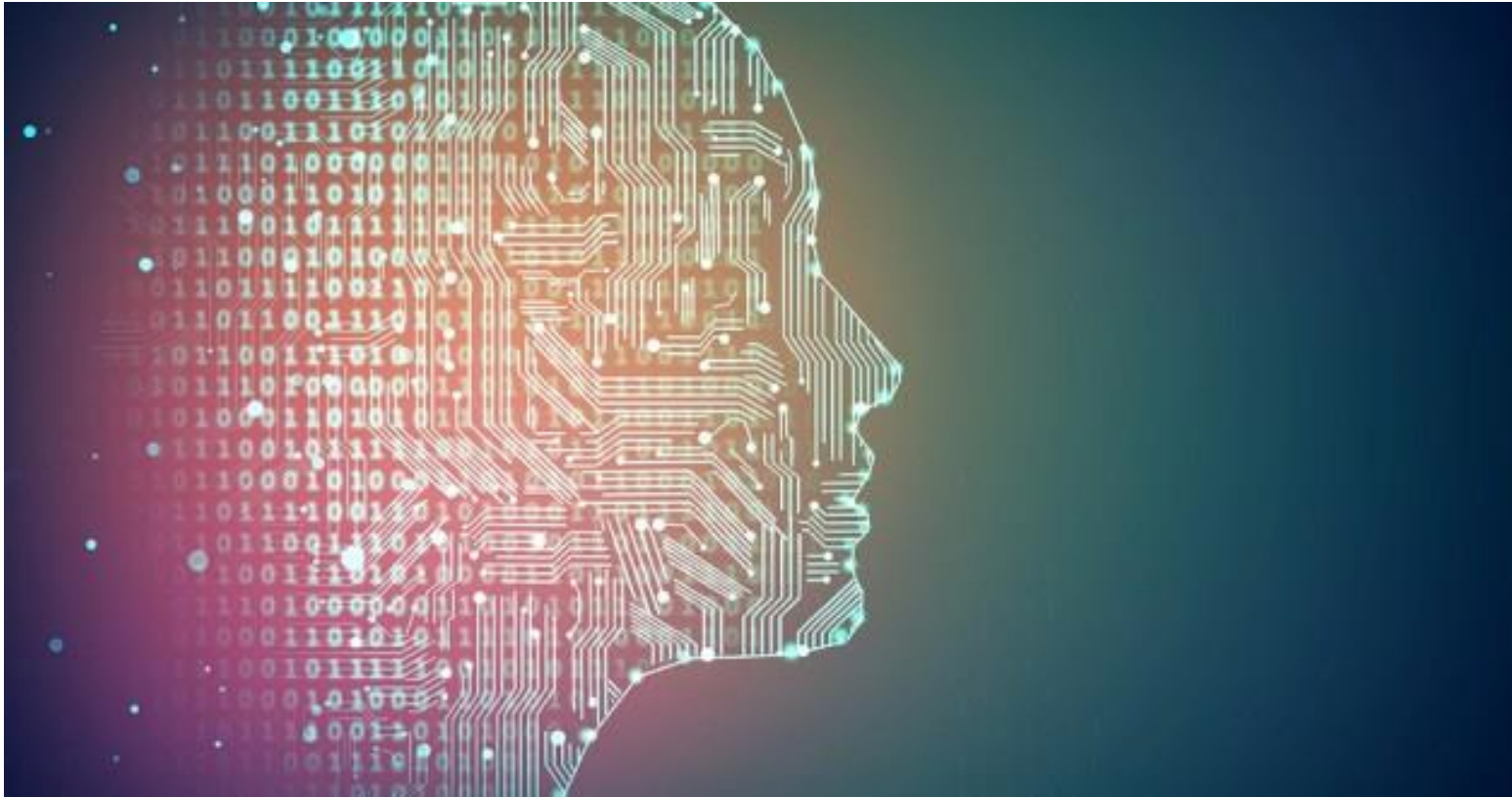


# **Generative AI: LLM: Vector Embeddings**

Thuan L Nguyen, PhD

## 2: Generative AI: LLM: Vector Embeddings



*AI Deep learning (Source: mindovermachines.com)*

# **3: Generative AI: LLM: Vector Embeddings**

1. Gen AI: LLM: Vector Embedding: Introduction & Definition
2. Gen AI: LLM: Vector Embedding: A Bit of History
3. Gen AI: LLM: Vector Embedding: Main Concepts
4. Gen AI: LLM: Vector Embedding: The Essence & Significance
5. Gen AI: LLM: Vector Embedding: Essential Parameters & Properties
6. Gen AI: LLM: Vector Embedding: Beyond Text Embeddings
7. Gen AI: LLM: Vector Embedding: Challenges & Issues

# 5: Generative AI: LLM: Vector Embeddings

## Artificial Intelligence: Generative AI

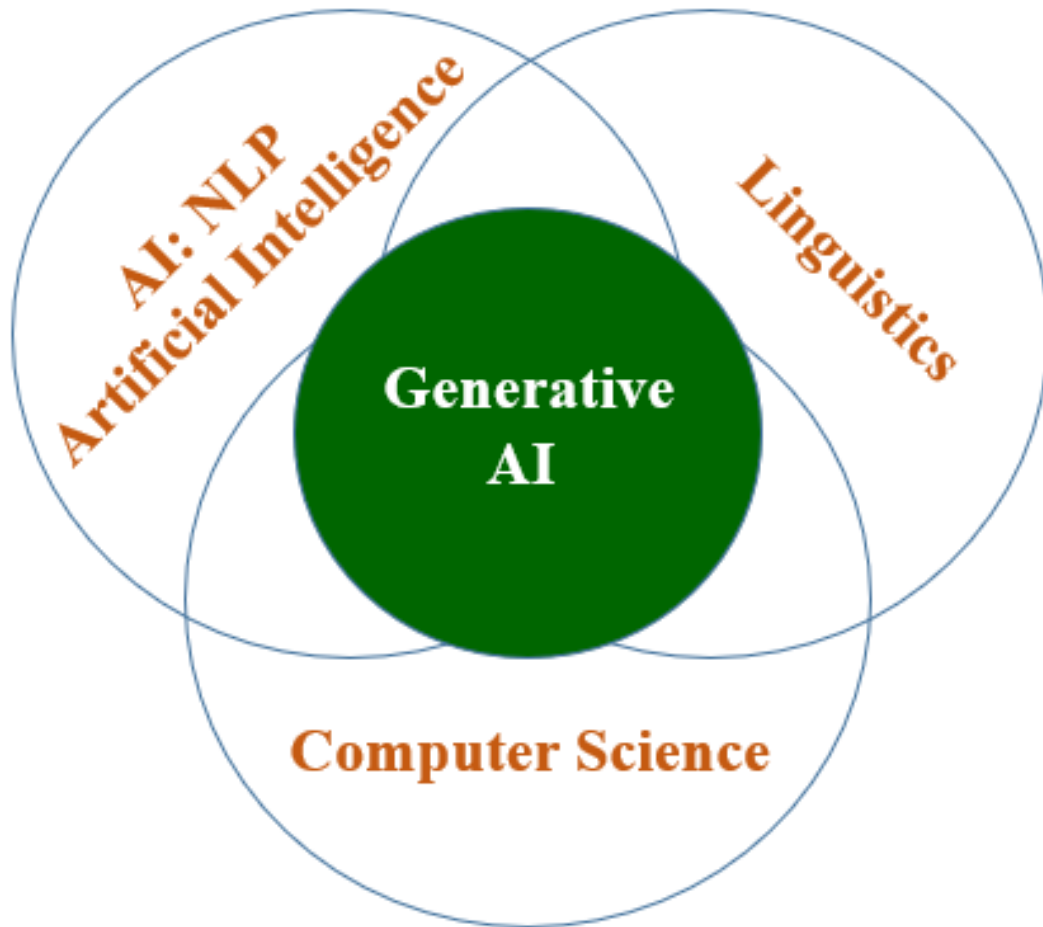
### What is It?

Generative AI: A category of artificial intelligence focused on using AI deep learning models to generate new contents, including text, images, audio, video, and more. The contents are novel but look realistic and may be indistinguishable from human-created ones.

# 5: Generative AI: LLM: Vector Embeddings

Artificial Intelligence: Generative AI: LLM

What is It?



**Generative AI** is based on the NLP technologies such as Natural Language Understanding (NLU) and Conversational AI (AI Dialogues) - Those among the most challenging tasks AI needs to solve.

# 6: Generative AI: LLM: Vector Embeddings

## Artificial Intelligence: Generative AI: LLMs

### Large Language Models

**Large Language Models (LLMs)** are **revolutionary AI Deep Learning neural networks** that excel in **natural language understanding (NLU)** and **content generation**.

- “LARGE” in LLMs refers to the vast scale of data and parameters used to train them, allowing LLMs to develop a comprehensive understanding of language.
- Being particularly transformer-based models trained on massive text datasets using deep learning techniques,
- Able to learn complex language patterns, capture nuances like grammar and tone, and generate coherent and contextually relevant text

# 7: Generative AI: LLM: Vector Embeddings

## Text Vector Embeddings: Overview

- In the field of Natural Language Processing (NLP), words in human language are symbolic, and computers could hardly directly grasp their meaning.
- Text vector embeddings provide a breakthrough solution by mapping words, sentences, or entire documents into a numerical space.
- This allows machines to process text semantically, leading to powerful applications like search, question-answering, and more.

# 8: Generative AI: LLM: Vector Embeddings

## Text Vector Embeddings: A Definition

- A dense, multi-dimensional numerical representation of a piece of text.
- Each number in this vector (or array) contributes to capturing various aspects of the text's meaning, from straightforward semantic relationships to subtle contextual shades.
- The dimensions (length) of these vectors often range from hundreds to even thousands of numbers.
- In summary, text vector embeddings are the numerical representations of words, phrases, sentences, or even entire documents that enable machines to grasp the essence and relationships inherent in human language.



# 9: Generative AI: LLM: Vector Embeddings

## Text Vector Embeddings: An Example

- Consider the words "dog", "cat", "run", and "play".
- It is possible to represent them in a highly reduced, two-dimensional embedding space where one dimension might relate to 'animal type' and the other to 'action':
  - dog: [0.8, 0.3]
  - cat: [0.9, 0.2]
  - run: [0.2, 0.7]
  - play: [0.4, 0.8]
- In this example of embedding:
  - "dog" and "cat" would be closer together due to their similarity in the 'animal type' dimension.
  - "run" and "play" would have greater proximity owing to the 'action' dimension.

# 10: Generative AI: LLM: Vector Embeddings

## Text Vector Embeddings: A Bit of History

- The concept of embedding words in vector spaces has a surprisingly long history. Its origins can be traced back to the **distributional hypothesis** proposed by **Zellig Harris** in the **1950s**.
  - In other words, the concept of representing words or concepts as **vectors** has its **roots** in the broader field of **distributional semantics**.
- This hypothesis proclaims that **words sharing similar contexts are likely to share similar meanings**.
  - The meaning of a word can be inferred from the contexts in which it frequently appears
    - For example: You shall know a word by the company it keeps.
- However, the concrete computational methods to realize word embeddings would take additional decades to materialize. It was in the **early 2010s** that **word embeddings like Word2Vec**, introduced by a team led by **Tomas Mikolov at Google**, gained widespread prominence.
  - These techniques **ignited a revolution in NLP**, empowering machines to capture a remarkably rich understanding of linguistic nuances.

# 11: Generative AI: LLM: Vector Embeddings

## Text Vector Embeddings: Main Concepts

### Vector Representation:

- At the heart of text vector embeddings lies the conversion of words or text segments into numerical vectors.
- These vectors are typically dense (most values are non-zero) and live in a high-dimensional space.
- For instance, a word might be represented by a 300-dimensional vector where each dimension holds a real number.

### Semantic Similarity:

- The magic of vector embeddings is that they capture semantic relationships.
- Words that are similar in meaning tend to have vectors that are closer to each other in the vector space.
- This can be measured using metrics like cosine similarity. For example, the vectors for "cat" and "dog" would likely be closer together than "cat" and "airplane".

### Contextual Awareness:

Advanced embedding models like those used by large language models (LLMs) go beyond single words. They can encode entire sentences or paragraphs, taking into account word order and overall context to generate more meaningful representations.

# 12: Generative AI: LLM: Vector Embeddings

## Text Vector Embeddings: The Essence and Significance

The **significance of vector embeddings** lies in their ability to **bridge** the gap between the **qualitative world** of human language and the **quantitative realm** that machines thrive upon with the following features:

- **Understanding Similarities:** Vector embeddings facilitate the understanding of semantic similarities between words or pieces of text. By calculating measures like cosine similarity, we can determine how closely related two words or texts are in meaning.
- **Crucial NLP Tasks:** **Embeddings** are required for many advanced NLP tasks. Large language models (LLMs) fundamentally rely on vector embeddings to:
  - **Question & Answer:** Find relevant answer snippets from documents based on the similarity of the embedding for the question to the embedding of text passages.
  - **Search:** Retrieve the most relevant documents from a **vast** corpus based on how closely their embeddings match the embedding of a search query.

# 13: Generative AI: LLM: Vector Embeddings

## Text Vector Embeddings: Essential Parameters and Properties

- **Vocabulary Size:** This refers to the number of unique words the embedding model is trained on. A larger vocabulary results in a more comprehensive representation.
- **Embedding Dimensionality:** This is the number of dimensions (numbers) in each vector. Higher dimensionality generally translates into the potential to capture more nuanced semantic distinctions. However, it comes with increased computational costs.
- **Training Data:** The quality and quantity of data used to train an embedding model have a profound impact on its quality. Diverse and domain-specific data often lead to embeddings better tailored to particular tasks.
- **Pre-trained Embeddings OR Embedding Techniques:** Models like Word2Vec, GloVe, and more recently, transformer-based sentence embeddings offer pre-trained embeddings accessible to developers. These save the intensive training process and can be adapted to various tasks.

# 14: Generative AI: LLM: Vector Embeddings

## Vector Embeddings: Beyond Text Embeddings

The power of **vector embeddings extends beyond text**. They're being actively used for representing:

- **Images**: Images can be embedded into vectors, enabling image similarity search, image classification, and more.
- **Audio**: Audio waveforms can be represented as embeddings for speech recognition, music classification, or speaker identification.
- **Molecular Structures**: Representing molecules in vector form has exciting applications in drug discovery and materials design.
- **User Behavior**: In recommender systems, user preferences and actions can be embedded, allowing for personalized recommendations.

# 15: Generative AI: LLM: Vector Embeddings

## Vector Embeddings: Challenges and Issues

While incredibly useful, **vector embeddings** also have **limitations** and raise **important questions**:

**Bias**: Embeddings are susceptible to reflecting biases present in the training data, potentially leading to discriminatory AI systems.

**Data Quality**: The effectiveness of embeddings heavily relies on the quality of the data they're trained on.

**Interpretability**: The high-dimensional nature of embeddings makes them difficult for humans to directly interpret.

# **8: Generative AI: LLM: Transformers**

Artificial Intelligence: Generative AI: LLMs

**Large Language Models**