Page 1: Introduction to Artificial Intelligence

Artificial Intelligence (AI) refers to the simulation of human intelligence in machines that are programmed to think like humans and mimic their actions. The term may also be applied to any machine that exhibits traits associated with a human mind such as learning and problem-solving. AI is widely used today in applications such as recommendation systems, voice assistants, fraud detection, and autonomous vehicles.

Machine Learning is a subset of AI that focuses on building systems that learn from data. Instead of being explicitly programmed, these systems improve their performance as they are exposed to more data.

Page 2: Natural Language Processing and Embeddings

Natural Language Processing (NLP) is a field of AI that gives machines the ability to read, understand, and derive meaning from human languages. One of the most important techniques in modern NLP is the use of embeddings. Embeddings convert words, sentences, or documents into numerical vectors that capture semantic meaning.

These vectors allow computers to compare text based on meaning rather than exact word matches. For example, the sentences 'How does a neural network learn?' and 'How are deep learning models trained?' are considered similar in embedding space, even though the wording is different.

Page 3: Retrieval-Augmented Generation (RAG)

Retrieval-Augmented Generation (RAG) is a technique that combines information retrieval with text generation. In a RAG system, relevant documents are first retrieved from a knowledge base using embeddings and similarity search. These retrieved documents are then provided as context to a large language model to generate accurate and grounded responses.

RAG is especially useful for question answering over private documents such as PDFs, manuals, or internal knowledge bases. By grounding responses in retrieved data, RAG systems reduce hallucinations and improve factual accuracy.