

Train a Custom AI Model on Your PDF Documents Using LangChain

This is a sample PDF document used for demonstrating PDF ingestion, chunking, and embedding creation with LangChain. The content here mimics real-world documents such as manuals, reports, or knowledge base articles.

Section 1: Introduction

LangChain enables developers to build applications powered by large language models using retrieval-augmented generation, vector stores, and document loaders.

Section 2: Why Use Custom Data?

Using your own documents allows AI models to answer domain-specific questions without requiring expensive fine-tuning. Instead, embeddings and vector search provide high-quality responses.

Section 3: Sample Knowledge

- LangChain supports multiple vector stores including FAISS and Pinecone.
- RecursiveCharacterTextSplitter is recommended for splitting PDF content.
- Embeddings convert text chunks into numeric vectors.

End of sample PDF.