Sample Company Knowledge Base

# 1. Introduction

This document contains sample structured and unstructured content to help you practice Word document parsing for Retrieval-Augmented Generation (RAG) pipelines.

# 2. Company Overview

Aithon AI is an artificial intelligence solutions company specializing in multimodal intelligence, autonomous systems, and applied deep learning.

Our mission is to enable human-AI collaboration at scale.

# 3. Key Products

• AithonVision — Real-time visual analytics platform for retail and logistics.

• AithonVoice — Speech-to-intent engine for multilingual environments.

• AithonRAG — Document intelligence system for enterprise knowledge retrieval.

# 4. Frequently Asked Questions (FAQs)

|  |  |
| --- | --- |
| Question | Answer |
| What is Aithon AI? | An AI solutions company focused on multimodal intelligence. |
| Where is the company located? | Headquartered in San Francisco, CA. |
| How many employees does the company have? | Approximately 200 worldwide. |
| Does Aithon AI offer APIs? | Yes, public and enterprise APIs are available. |

# 5. Technical Notes

The RAG pipeline typically involves three stages: chunking, embedding, and retrieval. Chunking strategies can use recursive character splitters, semantic chunking, or layout-aware parsing. Embedding can be performed using OpenAI embeddings, Sentence Transformers, or other vector encoders. Retrieval uses similarity search mechanisms like FAISS, Pinecone, or Weaviate.

In production, metadata like section titles, page numbers, and document hierarchy should be preserved for better retrieval performance.

# 6. References

1. https://aithon.ai/docs  
2. https://github.com/openai  
3. https://docs.langchain.com