Solar RAG Chatbot Test Document

This is a comprehensive test document for the Solar RAG chatbot system.

It contains multiple paragraphs and sections to test chunking functionality.

Introduction to RAG Systems:

Retrieval-Augmented Generation combines information retrieval with

language generation to provide more accurate and contextual responses.

Key Components:

1. Document Processing - Extract and prepare text from various sources

2. Text Chunking - Split documents into manageable pieces

3. Embedding Generation - Convert text chunks into vector representations

4. Vector Storage - Store embeddings in specialized databases

5. Retrieval and Generation - Search relevant chunks and generate responses

Solar AI Embedding Models:

Solar embedding models excel at Korean language processing and provide

high-quality semantic representations for various domains and use cases.

Testing Methodology:

This document is designed to test multiple aspects of the RAG system:

- Text extraction from PDF format

- Proper chunking of long content

- Embedding generation for each chunk

- Vector database storage and retrieval

Expected Behavior:

The system should split this document into multiple chunks based on

the configured chunk size and overlap parameters. Each chunk should

be converted to embeddings and stored in the Pinecone vector database.