

Vector Database and AI Integration

This is the first paragraph of the test document. It serves to test the "Paragraph Strategy". The code should identify the double newline after this block and treat this entire text as a single chunk. Vector databases are essential for storing embeddings efficiently.

This is the second paragraph. Here we are testing the "Sentence Strategy". This is the first sentence. This is the second sentence, which is slightly longer. And here is a third sentence to ensure the NLTK tokenizer is working correctly.

This is the third paragraph, designed to test the "Fixed Size Strategy". If you define a small chunk size (like 200 characters) in the JSON configuration, this paragraph might be split in the middle. We want to verify that the overlap mechanism works as expected so that no context is lost between the chunks.