Paper: BERT (sample notes)

BERT pretrains deep bidirectional transformers using masked language modeling.

It uses next sentence prediction during pretraining and then task-specific fine-tuning.

The approach significantly improved benchmarks in QA and language understanding tasks.

This sample file is included for quick local RAG testing.