PDF 1: Purpose of RAGFlow

RAGFlow is an open source Retrieval-Augmented Generation (RAG) engine designed to turn raw documents into reliable context for large language models. Its purpose is to make it practical to build an AI assistant that can answer questions using your private files, while staying grounded in the source material.

In a typical workflow, you ingest documents into a dataset, the system parses and cleans them, then chunks and indexes the content for search. At question time, RAGFlow retrieves the most relevant passages and sends them to the model as context. This reduces hallucinations and improves traceability because answers can be tied back to specific excerpts, often with citations.

RAGFlow focuses on deep document understanding, so it can handle complex formats beyond plain text, such as PDFs with tables, images, and mixed layouts. It provides an end-to-end pipeline: data ingestion, retrieval, chat, and agent style orchestration. It also exposes user interfaces and APIs so teams can experiment quickly, then operationalize the same flows in production. For a parsing test, the key idea is that the platform converts heterogeneous files into structured text chunks that are ready for retrieval, ranking, and quoting.