

Docling

Docling bundles PDF document conversion to JSON and Markdown in an easy, self-contained package.

Features

Converts any PDF document to JSON or Markdown format, stable and lightning fast.

Understands detailed page layout, reading order and recovers table structures.

Extracts metadata from the document, such as title, authors, references and language.

Includes OCR support for scanned PDFs.

Integrates easily with LLM app / RAG frameworks like LlamaIndex and LangChain

Provides a simple and convenient CLI.