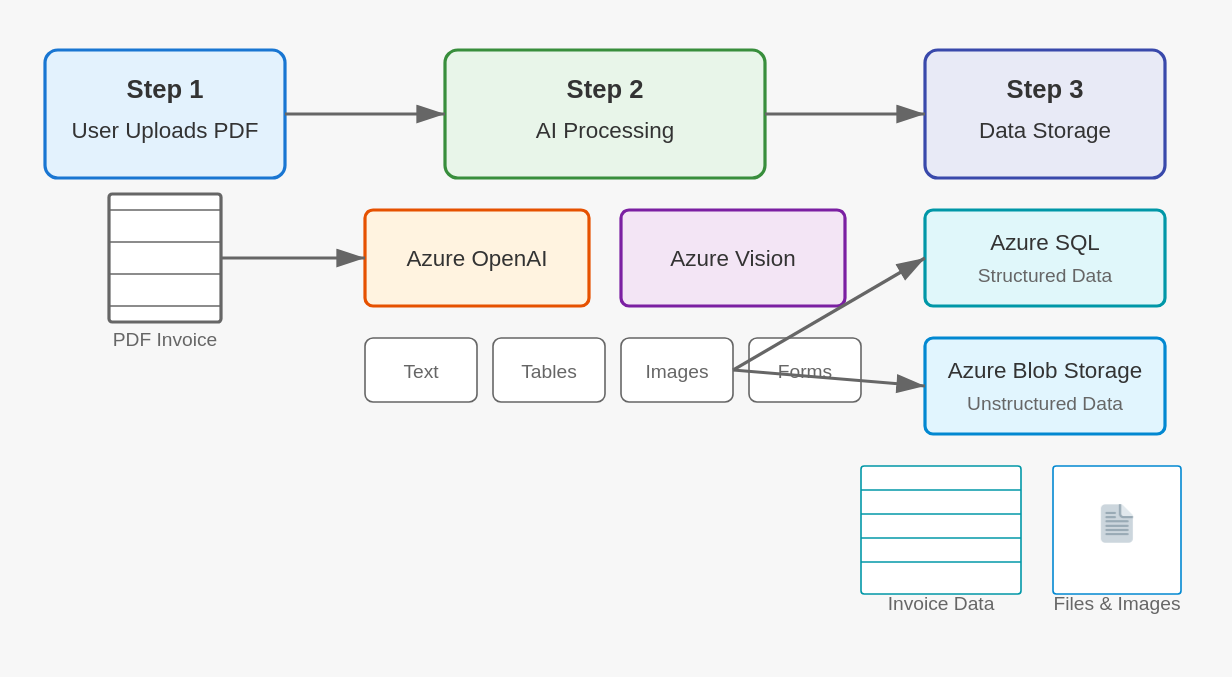
# **MPower Health: PDF Invoice Processing System**

****

**The block diagram illustrates a comprehensive PDF invoice processing system using Azure services. Let me explain each component in detail:**

**Step 1:** User Uploads PDF This is the initial stage where users submit their PDF invoices into the system. The diagram shows a PDF invoice flowing into this process block, representing the raw document that contains financial information that needs to be processed. This step serves as the entry point for all documents in the workflow.

**Step 2:** AI Processing This central component handles the intelligent analysis of the uploaded PDFs. It consists of two primary AI services:

**Azure OpenAI:** This service applies natural language processing to interpret and extract meaningful text content from the invoices. It can understand context, identify key information like invoice numbers, dates, and amounts, and transform unstructured text into structured data points.

**Azure Vision:** This image processing service analyzes the visual elements of the PDF. It can recognize and interpret graphics, logos, stamps, signatures, and other visual components that may be present in the invoice documents.

Content Categories (Text, Tables, Images, Forms) The AI processing step classifies the extracted content into four distinct categories:

* Text: Plain textual information like company details, descriptions, and terms
* Tables: Structured data presented in tabular format, such as line items or pricing details
* Images: Visual elements like company logos or product pictures
* Forms: Standardized document sections with specific layouts and fields

**Step 3:** Data Storage The final step involves storing the processed information in appropriate Azure storage solutions:

**Azure SQL (Structured Data)** This relational database stores well-organized, structured information extracted from the invoices. The "Invoice Data" box below shows how tabular information is systematically stored in database tables with defined schemas, making it easily queryable for reporting and analysis.

**Azure Blob Storage (Unstructured Data)** This object storage solution maintains the original files and images. The "Files & Images" box demonstrates how raw documents and visual elements are preserved in their native format, ensuring complete data retention while maintaining accessibility.

The entire workflow demonstrates a sophisticated document processing pipeline that transforms unstructured PDF invoices into organized, searchable data while preserving all original content in appropriate storage formats.