LedgerScribe: Invoice Parser & Journal Generator

Overview:

LedgerScribe is an AI-powered invoice processing and accounting automation tool that streamlines the manual tasks of extracting, analyzing, and recording financial data from vendor invoices. Built using Python, Streamlit, OpenAI's GPT models, and pandas, the application provides a smart, interactive interface for parsing PDF invoices and generating corresponding journal entries and inventory records in real-time.

Key Features:

- **Invoice Upload & OCR Extraction:**
  Users can upload PDF invoices, which are processed using PDFPlumber to extract raw text data. The extracted content is then fed to OpenAI's GPT-4 model to identify key invoice fields such as vendor name, invoice number, invoice date, line items, subtotal, taxes, and total amount.
- **Field Structuring & Validation:**
  GPT-generated data is normalized into structured formats (JSON, DataFrames), and enhanced with derived fields such as unit cost (calculated from amount and quantity) to ensure data usability and accuracy.
- **Journal Entry Generation:**
  Based on vendor or extracted invoice content, the system intelligently suggests journal entries. For example, an inventory-related purchase triggers a debit to “Inventory Purchases” and a credit to “Accounts Payable.” Users can edit these entries before saving.
- **Ledger Management:**
  Confirmed journal entries are stored in a structured ledger file with persistent serial numbering. This enables audit trails and accurate tracking of financial transactions.
- **Inventory Tab with Aggregation:**
  Inventory-related invoice line items are automatically synced to an file. The Inventory tab shows aggregated stock by item, including total quantity, total purchase cost, and average unit cost providing insights into procurement trends and valuation.

Technical Stack:

- **Frontend/UI:** Streamlit (custom tabbed interface with editable tables)
- **Backend/Logic:** Python (pandas, PDFPlumber, OpenAI API)
- **Storage:** Excel files (Ledger, Inventory) using pandas.to_excel
- **AI Integration:** GPT-4 for invoice field extraction and category suggestion

Use Cases:

- SMEs automating invoice entry and ledger maintenance
- Accountants or finance teams seeking AI assistance for bookkeeping
- Inventory managers tracking purchases and cost per unit

Impact:

LedgerScribe eliminates repetitive data entry, reduces human error in journal recording, and simplifies inventory reconciliation. It bridges the gap between raw invoice data and structured accounting output, offering an end-to-end intelligent automation solution for financial document processing.