AgentAnalyzer

AgentAnalyzer is built on an agent-based architecture, where multiple independent yet interconnected modules work together seamlessly. It breaks down complex data transformation tasks into smaller, specialized processes, each handled by a dedicated agent. The main objective is to improve efficiency, accuracy, and adaptability in data processing using AI-driven intelligence.

The Agentic Workflow: Step-by-Step Methodology

1. **Data Extraction (Module: extract_csv.py)**

Objective: Retrieve raw data from CSV sources.

Mechanism:

- Load CSV data into a Pandas DataFrame, standardizing formats.
- Handle encoding issues, missing values

2. **Merge Strategy & Data Merging (Modules: merge_strategy.py & merge_csv.py)**

Objective: Seamlessly integrate multiple data sources into a unified dataset using a strategy.

Mechanism:

- Employs OPENAI to determine the best strategy to merge the data
- Use merge_strategy determined from previous step to combine records and create a master dataset.
- Detects and resolves duplicate entries while preserving accuracy by grouping the data based on key column.

Challenges: Managing data conflicts, duplicate handling, and format discrepancies.

3. **Data Population & Structuring (Module: populate_template.py)**

Objective: Populate cleaned and structured data into a standardized Excel template.

Mechanism:

- Load an Excel template using openpyxl.
- Determine the data mapping between the template topics and master databse.
- Perform transformations on data through OPENAI API call to match with the template.
- Insert transformed data into designated fields keeping the format consistent with the template.
- Generate a finalized, structured output file.

Challenges: Preventing formatting inconsistencies and ensuring full adherence to template specifications.

Limitation

- Not able to analyze template files which have merged cells.