Methodology Document
Financial AI Agent

1 System Overview

This project implements an end-to-end **Financial AI Agent** designed to automate the extraction, validation, analysis, and interpretation of financial information from Latvian annual report PDFs. The system integrates:

- Large Language Models (LLMs)
- Schema-validated structured extraction
- Financial ratio computation
- Cross-company visualization
- LLM-generated peer review
- A user-friendly Gradio interface

The pipeline consists of:

1. Document ingestion and preprocessing
2. LLM-based structured extraction
3. Financial ratio computation
4. Aggregation and visualization
5. Peer review generation
6. Interface deployment

2 Data Ingestion and Preprocessing

2.1 PDF Loading

PDFs are processed using `PyPDFLoader`. For each uploaded document:

- The first eight pages are extracted (balance sheet and profit & loss).

- Pages are concatenated into a unified text corpus.

2.2 Unicode and Text Sanitization

To ensure cross-platform stability:

- Unicode characters above `0xFFFF` (emojis, astral symbols) are removed.

- Latvian characters (ā, ē, č, ģ, ļ, ņ, š, ž) are preserved.

  This prevents encoding errors while maintaining linguistic accuracy.

3 Structured Financial Information Extraction

3.1 Pydantic Schema Design

A custom `Pydantic` model (`LatvianFinancials`) enforces strict validation for:

- Company identifiers

- Profit & Loss variables

- Balance sheet variables

- Optional fields for missing data

3.2 LLM-Based Parsing with LangChain

Extraction uses:

- `PromptTemplate`

- `PydanticOutputParser`

- Selectable LLM backend (GPT-4o, Llama-3, Gemini)

The prompt enforces accounting rules:

- Remove thousand separators

- Convert bracketed values to negatives

- Prefer reporting-year values

- Use Latvian row codes ("Rindas kods")

This produces validated, machine-readable financial data.

4 Financial Ratio Computation

Ratios are computed using extracted values.

4.1 Liquidity

Current Ratio

$$\text{Current Ratio} = \frac{\text{Current Assets}}{\text{Current Liabilities}}$$

4.2 Profitability

Net Profit Margin (%)

$$\text{Net Margin} = \frac{\text{Net Profit}}{\text{Net Turnover}} \times 100$$

4.3 Solvency

Debt-to-Equity Ratio

$$\text{Debt/Equity} = \frac{\text{Short-term Debt} + \text{Long-term Debt}}{\text{Equity}}$$

Fallback logic handles missing or zero values.

5 Aggregation and Visualization

5.1 Data Aggregation

Outputs from multiple companies are combined into a `Pandas DataFrame` for:

- Cross-company comparison

- Year-aligned analysis

5.2 Interactive Visualization

Using Plotly Express, the system generates:

- Grouped bar charts for liquidity, profitability, solvency

- Interactive hover and labels

6 LLM-Generated Peer Review Analysis

The system generates a qualitative analysis by:

- Passing the aggregated DataFrame as a markdown table

- Evaluating liquidity, profitability, leverage

- Producing a structured narrative and recommendation

Output language is selectable (English or Latvian).

7 User Interface and Deployment

Built using Gradio Blocks, the interface supports:

- API key input

- LLM provider selection

- Language selection

- Multi-PDF upload

- Display of:

  - Structured financial table

  - Interactive charts

  - Peer review text

8 Methodological Advantages

This methodology:

- Combines accounting logic with LLM reasoning

- Ensures schema-validated outputs

- Supports multi-model robustness

- Bridges quantitative metrics and qualitative interpretation

The Financial AI Agent provides a scalable, explainable solution for automated Baltic financial report analysis.