

# HelpBot v2.3 [Internal Policy Update] - Documentation

## Executive Summary

The FINREP Extractor is a custom-built AI system designed for the automated extraction of key financial metrics and data points from uploaded financial reports. It is embedded within Fitch's internal suite of analytics tools, accessible via API and Web UI for internal users. Its primary function is to streamline data acquisition for internal analysis, improving efficiency and accuracy in financial reporting workflows.

## Business Requirements

Utilizes document parsing, OCR, and NLP for metric extraction. Based on information extraction theory and deep learning approaches

## Business Background and Rationale

**Use Case:** The system supports internal financial analysis by automating the extraction of key metrics from financial reports, reducing manual data entry and potential errors.

**Users:** Internal Fitch analytics teams and business units requiring financial data for analysis.

**New/Existing System:** enhancement

## Applicable Policies, Standards, and Procedures

- Copyright laws
- Fitch Data Governance Policy

## Functional Requirements

- API integration capabilities
- AWS Redshift Access, Atlas Access via REST API

- CSV data exports
- Hosted on internal servers; API/Web UI endpoints
- Outputs data in JSON/CSV
- RESTful API endpoints
- The FINREP Extractor is a custom-built AI system designed for the automated extraction of key financial metrics and data points from uploaded financial reports. It is embedded within Fitch's internal suite of analytics tools, accessible via API and Web UI for internal users. Its primary function is to streamline data acquisition for internal analysis, improving efficiency and accuracy in financial reporting workflows.
- Outputs data in CSV data exports, Fitch Data Governance Policy, Historical portfolio data, Portfolio and market data
- Access control for internal users only

## Development Dataset

**Overview:** Pre-approved reports from internal repositories.

**Data Sources and Extraction Process:** Reports sourced from CSV data exports, Fitch Data Governance Policy, Historical portfolio data, Internal Fitch repositories, Portfolio and market data transformed using processing pipelines.

**Vendor Data/Data Proxies:** no-vendor

**Data Sampling:** Achieved >95% extraction accuracy on training set.

**Data Quality:** Risks: Extraction errors, data quality issues, unauthorized access  
Metrics: Extraction accuracy, input format validation, access logs Review: Monthly dashboard; integrated with internal monitoring tools

## Methodology, Theory and Approach

**Description:** Improved OCR preprocessing for scanned documents

**Limitations and Risks:** Plan to expand to additional report types and automate error correction

## System Calibration

**Development Code:** [github.com/fitch-field/helpbotmodel](https://github.com/fitch-field/helpbotmodel)

**Key System Assumptions:** All documents are genuine, pre-approved financial reports; no personal data present.

## Developer Testing

**In-Sample Back Testing Analysis:** Reports sampled by year and sector for training/testing. **Out-of-Sample Back Testing Analysis:** Stress tested with batch uploads and large files. **In-Sample Back Testing Analysis:** Achieved >95% extraction accuracy on training set. **Out-of-Sample Back Testing Analysis:** Maintained >92% accuracy on test set. **Benchmarking/Challenger Tool Analyses:** Outperformed manual extraction in speed and accuracy. **Out-of-Sample Back Testing Analysis:** Tested on reports with missing headers, scanned images; robust except for very poor scans.

## Governance

**Ethical Considerations:** - **Fairness:** No risk of discrimination; only financial data processed. - **Safety:** No personal data; complies with internal and external regulations. - **Security:** Restricted to internal access; Access logs and security events, Authentication systems, Input validation to prevent manipulation, Unauthorized access. - **Robustness:** Output accuracy monitored; retraining scheduled annually. - **Explainability:** Processing steps logged and reviewable by analysts. - **Transparency:** System functionality documented for users. - **Governance:** Roles assigned per organizational AI Governance Guidance.

## Risk Monitoring Plan

**Risks:** Processing errors, Data quality issues, Unauthorized access

**Metrics:** Processing accuracy, Input format validation, Access logs

**Review:** Monthly dashboard; integrated with internal monitoring tools

## Lessons Learned and Future Enhancements

- enhancement
- Improved OCR preprocessing for scanned documents
- Plan to expand to additional report types and automate error correction

## Deployment Specification

**Technical Requirements:** Hosted on internal servers; API/Web UI endpoints

**Architecture Diagram:** [Insert data flow architecture]

**Process Flow Diagram:** [Insert workflow diagram]

**Engineering Interface:** The FINREP Extractor is a custom-built AI system designed for the automated extraction of key financial metrics and data points from uploaded financial reports. It is embedded within Fitch's internal suite of analytics tools, accessible via API and Web UI for internal users. Its primary function is to streamline data acquisition for internal analysis, improving efficiency and accuracy in financial reporting workflows.

**Implementation Code:** Repository at [internal location]

**Production and Testing Environment Access:** Access via internal roles

**Upstream and Downstream Models/Applications/Dependencies:** Upstream: internal repositories; Downstream: analytics dashboards

**User Acceptance Testing ('UAT'):** UAT completed; summary available in documentation

**Retention and Back Up:** Custom retention policy for processed data; backups at [internal location]

**User Guides (if applicable):** Step-by-step guide attached

**Other:** Data dictionary and technical specs attached

## Attachments

- images.png (8.8 K) - govern-metadata/8c1e4559-549c-42e8-a05a-760c99665349/b3d07aa3-076b-4313-971d-4ccfea2c27ab/images.png

## Visual Governance Dashboard

Auto-generated from recent trace data (or representative sample if no credentials were present).

AI Governance & Risk Management Dashboard

Real-time Monitoring & Compliance Assessment

