AI-Powered Tax Calculation and ITR Suggestion System

Moin Mulla[1], Sahil Mulani[2], Krutika Kamble[3], Dr. S.B. Takmare[4], Dr. S. V. Patil[5]
[1,2,3]UG – CSE (AI & ML) Engineering, Shivaji University, Kolhapur, Maharashtra
[4] Associate Professor, CSE (AI & ML) Engineering, DYPCET, Kolhapur, Maharashtra
[5] HOD, CSE (AI & ML) Engineering, DYPCET, Kolhapur, Maharashtra
Emails: mullamoin31@gmail.com[1], sahilm8940@gmail.com[2], krutikakamble@gmail.com[3],
sachintakmare85@gmail.com[4] , sidheshwar.patil@gmail.com[5]

Abstract

Filing Income Tax Returns (ITR) in India can be confusing and time-consuming due to frequent changes in tax rules and the need to manually interpret financial data. Most accounting tools can store records but do not help users understand which tax rules apply, which ITR form to choose, or how their final tax amount is calculated. This paper presents an AI-Powered Tax Calculation and ITR Suggestion System that combines Retrieval-Augmented Generation (RAG) and Large Language Models (LLMs) to make tax filing simpler and more transparent. The system reads financial data directly from Tally XML files, analyzes income and expenses, and applies the latest tax regulations using a hybrid approach that blends rule-based logic with intelligent knowledge retrieval. It also generates clear, human-readable explanations describing tax liability, deductions, and the recommended ITR form. Testing with datasets showed accurate results for individuals, firms, and companies. The proposed system reduces manual effort, ensures up-to-date compliance, and improves user confidence through easy-to-understand AI-generated reports.
Keywords: Tax Computation, ITR Suggestion, RAG, LLM, Explainable AI.

1. Introduction

Tax compliance is becoming increasingly complex due to constant changes in regulations, policies, and different formats of financial reporting. The current tax software uses static rule sets and manual interpretation, which are prone to errors and not scalable for SMEs and service providers. Recent advances in Artificial Intelligence (AI), especially Large Language Models (LLMs) and Retrieval-Augmented Generation (RAG), allow for automated reasoning about regulatory text, with answers grounded in external knowledge. This makes AI amenable to intelligent tax calculation. But current AI and rule-based systems require manual updates to knowledge bases. Because tax regulations and ITR guidelines keep changing from time to time, there is a need for self-updating AI systems in the regulatory space. The proposed research work is about an AI-Powered Tax Calculation and ITR Suggestion System, which extracts financial information from Tally XML files and combines an Auto RAG Updater with LLM Reasoning. The system is designed to fetch the latest tax regulations from the government website and embed them in a vector database to ensure regulation-compliant tax calculations. The main contribution of this research work is the self-updating RAG system that does not rely on static rule sets. [1]

1.1. Challenges in Tax Computation Systems

Tax computation is a complex process that includes a series of rules, deductions, and changing slabs. Rule-based systems lack the ability to update automatically, and hierarchical data in Tally is hard to understand for intelligent processing. [2]

1.2. Need for Self-Updating AI Systems

AI systems should be able to update knowledge automatically in dynamic regulations. Retrieval mechanisms along with LLM reasoning capabilities will provide error-free and regulation-compliant tax support. [3]

2. Methodology

This section outlines the architectural design and functional workflow of the proposed AI-driven tax calculation and compliance system. The system implements a systematic processing workflow that commences with XML upload and follows through with financial data extraction, tax calculation,

regulation retrieval, and explainable reporting. The system architecture is depicted in Fig. 1. The proposed framework is organized into five major phases:

- XML Upload and Preprocessing
- Financial Data Extraction [4]
- Tax Computation Pipeline
- RAG-Driven Regulatory Retrieval
- Report Generation and Compliance Validation
- Report Generation and Storage Figure 1 shows Methodology

Figure 1 Methodology

2.1.XML Upload and Preprocessing

The system supports the upload of XML files generated by Tally ERP software with financial data. The uploaded XML file is validated for correctness. A hierarchical parsing module is employed to parse hierarchical ledger components and eliminate structural errors. [5]

2.2.Financial Data Extraction

A hierarchical data extraction engine traverses the complex ledger and account relationships in the XML file to extract tax-related data like gross receipts, expenses, depreciation, assets, and liabilities. The extracted data is then formatted into a standardized schema, which serves as the input for tax calculation and ITR identification. [6]

2.3.Tax Computation Pipeline

The Tax Calculation Engine adopts a hybrid two-layer model:

- Rule-Based Layer: Engages in deterministic

calculations like income aggregation, deduction, depreciation, and application of tax slabs, surcharge, and cess.
- LLM-Assisted Reasoning Layer: Deals with situations involving ambiguous and context-dependent tax provisions that cannot be completely codified through static rules. The LLM facilitates the interpretation of complex regulatory conditions in the computation process. [7]

2.4.RAG-Driven Regulatory Retrieval

For the purpose of tax regulation compliance, the system combines a Retrieval-Augmented Generation (RAG) pipeline with an auto-update feature:

- A web scraping component periodically fetches tax circulars, notifications, and policies from government websites.
- The scraped documents are preprocessed for segmentation and embedding in a ChromaDB vector database.
- When performing tax calculations, the system retrieves relevant regulatory content based on semantic similarity search to anchor the LLM's reasoning process to the latest tax policies. [8]
- The Auto RAG Updater feature automatically updates the system without human intervention, ensuring that the tax calculation process is always up-to-date with the latest tax policies.
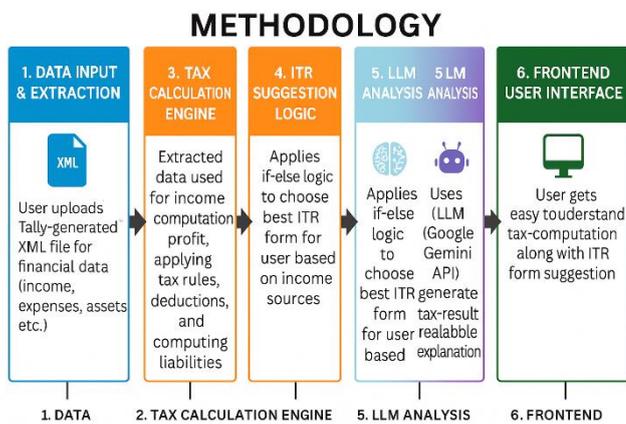
2.5.LLM-Based Explanation and ITR Recommendation

The LLM explanation component produces human-friendly outputs such as the tax rules applied, deduction statements, regulatory citations, and recommended ITR forms based on the extracted financial information. [9]

2.6.Report Generation and Storage

The system provides a detailed tax report on calculations, explanations, and ITR suggestions. All processed data is stored in MongoDB Atlas.

3. Results and Discussion

The proposed AI-driven Tax Calculation and ITR Suggestion System was tested for its qualitative performance on various major components of the system, such as XML-based financial data extraction,

dynamic regulatory extraction, LLM-based reasoning, and overall tax calculation. Since the system is still in its nascent stage of development as an intelligent compliance solution, the testing and discussion of results are more focused on correctness, robustness, and explainability of the system rather than its numerical performance. [10]

3.1. Results

This section highlights the performance of the system on various major functional aspects.

3.1.1. Performance of XML Financial Data Extraction

The XML processing module was able to function well with different XML files produced by Tally ERP. The system was able to process the complex ledger structure, voucher, and financial fields related to taxes without any issues, despite the differences in the format and names of the ledgers. The extracted data was formatted in a standardized internal representation, which made it easy to integrate with the tax calculation engine. The stability of the extraction process is an indication that it is robust and can be used for automated financial data extraction.

3.1.2. End-to-End System Behavior

The entire processing chain, ranging from the XML file import to tax liability calculation and ITR suggestion, was fully automated without human intervention. The calculated tax amounts matched expertly validated calculations for both old and new taxation systems. The processing chain handled dynamic financial inputs and the latest regulatory changes obtained from the RAG pipeline. The final tax calculation and deduction processing is shown in Figure 2.
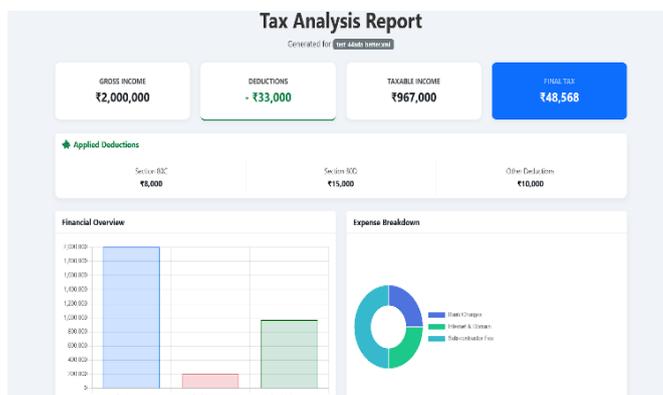

Figure 2 Tax Calculation and Deduction

3.1.3. User Experience Evaluation

The system provided explanations for users regarding the applied tax rules, deductions, and ITR form suggestions. The explainability component enabled users to interpret how the tax liabilities were calculated. The final output screen displaying tax summaries and ITR suggestions is shown in Figure 3.

Figure 3 Tax Summaries and ITR Suggestions

3.2. Discussion

The outcome emphasizes the feasibility of integrating financial data processing with adaptive AI-driven regulatory reasoning. The uniform performance of the XML extraction module validates the fact that ERP-based XML data is a robust platform for automated tax calculation, offering greater accuracy than document- or OCR-based systems. The hybrid tax calculation model, which combines deterministic rule-based reasoning with LLM-driven regulatory reasoning, was successful in processing standard and context-specific tax sections. The dynamic regulatory retrieval system is one of the most important aspects of the proposed system, which ensures that the

system is updated according to the latest government notifications and circulars. Unlike traditional static tax software, the Auto RAG Updater system eliminates the risk of knowledge obsolescence by regularly updating the regulatory content used for tax computation. The role of LLM-driven explanations in the proposed system cannot be overstated, as it is one of the most important aspects of the system that helps to establish user trust In conclusion, the system has great potential for use by small businesses, freelancers, and IT service providers who require accurate, up-to-date, and interpretable tax computation without having to depend much on tax professionals [11]

Conclusion

This paper has demonstrated an AI-assisted Tax Calculation and ITR Suggestion System that combines automated financial data processing, a dynamically updated Retrieval-Augmented Generation (RAG) pipeline, and Large Language Model (LLM)-based reasoning to enable accurate, interpretable, and regulation-compliant tax calculation. By overcoming the need to rely on static rule sets and facilitating real-time access to the latest regulatory content, the proposed system overcomes important challenges in traditional tax software.

Experimental results and analysis have shown the effectiveness of the system in handling XML-based financial data, retrieving the latest tax regulations, calculating tax liability, and providing interpretable results. The integration of data parsing and LLM-based reasoning enhances the interpretability and usability of the system, making it applicable to freelancers, IT service companies, and small to medium-sized enterprises. In conclusion, the proposed framework provides a scalable and future-proof solution for automated tax compliance in a dynamically changing regulatory landscape. Future research can be directed towards incorporating more sources of regulatory content, providing multilingual explanations, direct e-filing system integration, and enhanced financial reporting automation.

Acknowledgements

References

[1]. J. Luketina, A. Benkel, and C. G. Schuetz, "Using large language models for legal decision-making in Austrian value-added tax law: An experimental study," arXiv preprint arXiv:2507.08468v1, 2025.

[2]. N. Malali, "The role of retrieval-augmented generation (RAG) in financial document processing: Automating compliance and reporting," International Journal of Management Technology, vol. 12, no. 3, pp. 26–46, 2025.

[3]. K. Mora, "An analytical study on adoption and challenges of online ITR (Income Tax Return) filing by individuals," TIJER – International Research Journal, vol. 12, no. 7, pp. a297–a301, 2025.

[4]. X. Wang et al., "FinSage: A multi-aspect RAG system for financial filings question answering," arXiv preprint arXiv:2504.14493v2, 2025.

[5]. S. Sekhri, "Artificial intelligence and tax compliance: A paradigm shift," Chartered Secretary, pp. 93–97, 2025.

[6]. J. A. Gidisu, M. Celestin, M. K. Lumor, W. T. Dometi, and A. T. Amezuwey, "The role of artificial intelligence in tax compliance: Can AI reduce tax evasion and improve revenue collection?" Brainae Journal of Business, Sciences and Technology, vol. 9, no. 4, pp. 577–588, 2025.

[7]. V. Perot et al., "LMDX: Language model-based document information extraction and localization," arXiv preprint arXiv:2309.10952v2, 2024.

[8]. J. Wang, W. Ding, and X. Zhu, "Financial analysis: Intelligent financial data analysis system based on LLM-RAG," 2025.

[9]. N. Bhatt, "The impact of artificial intelligence on taxation: Opportunities, challenges, and future directions," International Journal of Advanced Legal Research, vol. 4, no. 3, 2024.

[10]. A. Bajpai, "Evaluating the impact of artificial intelligence on enhancing tax compliance and financial regulation," International Journal of Multidisciplinary Research and Technology, vol. 5, no. 6.

[11]. B. R. Varshini, "Use of artificial intelligence in scrutiny selection of income tax returns in tax administration," Indian Journal of Integrated Research in Law, vol. 4, no. 2, pp. 203–211.