International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 10 Issue: 10 | Oct - 2026 SJIF Rating: 8.659 ISSN: 2582-3930

TaxWise: A Deterministic and AI-Assisted Framework for Personalized Income Tax Preparation

Diganth S
Computer Science and Engineering
East West Institute of Technology
Bengaluru, India
diganth200515@gmail.com

Gowtham L
Computer Science and Engineering
East West Institute of Technology
Bengaluru, India
gowthamlokesh012@gmail.com

Ajay Gowda S
Computer Science and Engineering
East West Institute of Technology
Bengaluru, India
ajaygowda.s2005@gmail.com

Druva K
Computer Science and Engineering
East West Institute of Technology
Bengaluru, India
druva552005@gmail.com

Prof. Navyashree J
Computer Science and Engineering
East West Institute of Technology
Bengaluru, India
navyashree.j@ewit.edu.in

***Abstract*—Preparing an income tax return involves more than calculating tax. A taxpayer has to identify relevant income sources, understand deductions, select an appropriate return form, provide supporting documents, and interpret the final result. This paper presents a web-based tax preparation framework for Indian individual taxpayers that combines deterministic tax computation with conversational assistance and document-based information extraction. The system maintains a structured taxpayer profile and uses rule-based services for tax calculation and return selection, while the AI assistant is used for natural-language interaction and explanation. Assessment Year 2026–27 is currently supported, with return selection covering ITR-1 to ITR-4 within the implemented scope and explicit handling of unsupported conditions. The system also includes assessment-year-specific knowledge retrieval for the conversational layer. Uploaded documents are processed to produce candidate information that must be reviewed before being added to the taxpayer profile. The implementation uses Next.js and TypeScript on the frontend and FastAPI, Python, SQLAlchemy, and PostgreSQL on the backend. The design keeps numerical tax decisions separate from language generation so that calculations remain reproducible and easier to verify.**

***Index Terms*—income tax, tax preparation, deterministic computation, conversational AI, document intelligence, retrieval, ITR, India**

I. Introduction

Income tax preparation requires users to provide personal information, income details, deductions, tax payments, and supporting documents while also satisfying conditions associated with different return forms. For users who are unfamiliar with tax terminology, the process can become difficult even when online filing facilities are available.

Artificial intelligence has been increasingly explored in taxation for areas such as tax assistance, document processing, compliance, fraud detection, and knowledge-based advisory systems. Recent systems such as TaxCraft have investigated document processing, retrieval, and language models for personalized tax assistance [1]. Other studies have examined the use of artificial intelligence in GST-related processes and input tax credit reconciliation [2], [3]. Machine-learning methods have also been applied to tax fraud and default detection [4], [5]. Research on knowledge-grounded tax reasoning further shows the importance of retrieval and domain-specific information when language models are applied to tax-related tasks [6].

These approaches demonstrate the usefulness of AI, but tax preparation also contains calculations and eligibility decisions where reproducibility is important. A language model may provide a fluent response without necessarily producing the correct numerical result. The proposed system therefore separates these responsibilities.

The tax engine is responsible for numerical calculations, while the return selection service applies explicit rules to determine the applicable return form. The conversational component assists the user and explains verified results. A retrieval layer provides assessment-year-specific information to support contextual responses.

The main contributions of the project are:

- a deterministic tax calculation pipeline for supported Indian individual-tax scenarios;
- rule-based selection and preparation support for ITR-1 to ITR-4 within the implemented scope;
- conversational onboarding using the same structured taxpayer profile as the main application;
- document-assisted extraction with user confirmation before profile updates;
- assessment-year-specific retrieval for tax-related explanations;
- separation of numerical tax decisions from AI-generated language.

II. Related Work

Existing work on AI in taxation can be broadly grouped into tax advisory systems, tax administration applications, and tax-focused language models.

Tax advisory systems have explored the combination of document processing, retrieval, and language models. TaxCraft, for example, uses machine learning and language technologies to support personalized financial and tax assistance [1]. Such systems show how conversational interfaces can reduce the difficulty of finding and interpreting tax information.

Another area concerns tax administration. Research involving GSTN has examined the role of artificial intelligence in tax computation, compliance, return processing, and related administrative activities [2]. Similarly, research on input tax

credit reconciliation investigates how AI can reduce manual matching and validation work [3].

Machine-learning techniques have also been applied to identify suspicious tax behavior and predict tax defaults. Fraud-detection approaches focus on identifying unusual or potentially fraudulent activity rather than preparing an individual taxpayer's return [4]. The TAB-based approach similarly investigates machine learning for tax prediction and planning [5].

More recent tax-specific language-model research has emphasized knowledge grounding and domain adaptation. LLaMA-UTP, for example, combines domain-specific knowledge and language-model techniques for uncertain tax position analysis [6].

The present system does not attempt to reproduce these approaches in full. Instead, the literature motivates a practical separation of responsibilities: deterministic software handles calculations and eligibility, while AI and retrieval are used to improve interaction, explanation, and access to relevant tax information.

III. System Architecture

The system follows a layered client-server architecture. The frontend provides the user interface for profile management, conversational onboarding, document processing, tax calculation, and return preparation. The backend provides authenticated APIs and contains the tax engine, return-selection logic, document services, retrieval, and AI integration.

Figure 1 shows the main flow.

The architecture separates the calculation path from the conversational path. The deterministic services remain responsible for numerical tax values and return-selection decisions. The retrieval layer provides assessment-year-specific information to the assistant, while the language model converts verified information into natural-language responses.

This separation reduces the possibility that a generated response will silently alter the taxpayer's calculated result.

A. Frontend Layer

The frontend is implemented using Next.js, TypeScript, Tailwind CSS, and Zustand. The main workflow is centered around the taxpayer profile. Users can enter personal information, employment details, income, deductions, tax payments, bank information, and supporting documents.

The interface also provides conversational assistance alongside the structured profile. This allows the user to continue working with the tax profile while asking questions about the information being entered.

B. Backend Layer

The backend is implemented using FastAPI and Python. It exposes authenticated APIs for profile management, chat, documents, tax calculation, and return preparation.

SQLAlchemy is used for database interaction and PostgreSQL provides persistent storage. The separation of backend services allows the calculation and eligibility modules to operate independently from the AI provider.

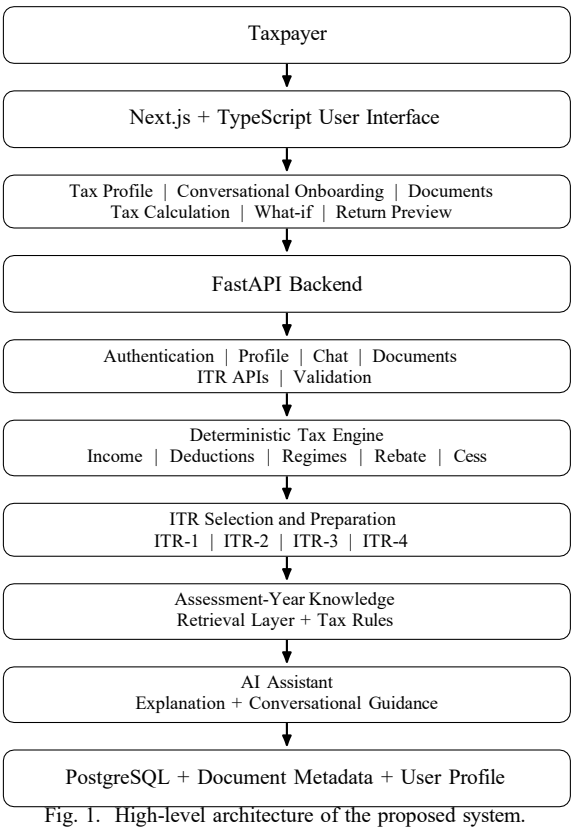


Fig. 1. High-level architecture of the proposed system.

C. Tax Profile and Data Isolation

The Tax Profile acts as the central representation of taxpayer information. It includes the assessment year, residential status, employment information, salary or pension income, house-property income, other income, capital gains, business or professional income, deductions, tax payments, bank details, and relevant return-selection information.

Authenticated operations are associated with the user's identity. Documents and tax profiles are isolated by the authenticated user. Sensitive values such as PAN and bank account information are masked where appropriate in preparation outputs.

IV. Deterministic Tax Computation

The tax engine is the numerical authority of the system. It accepts a validated taxpayer profile and selected tax regime and produces a structured calculation result.

Financial values are represented using Python's `Decimal` type. This avoids unnecessary floating-point rounding problems in monetary calculations.

A simplified representation of taxable income is:

$$Y = \max(0, G - D_s - D) \quad (1)$$

where G represents gross income, D_s represents the applicable standard deduction, and D represents other eligible deductions.

The tax before final adjustments can be represented as:

$$T_b = T_s + T_c \tag{2}$$

where T_s is the applicable slab-based tax and T_c represents supported special-rate tax such as applicable capital-gains tax.

After rebate, surcharge, and cess, the final liability can be expressed generally as:

$$T = \text{round}\,(\max(0, T_b - R) + S + C) \tag{3}$$

where R is rebate, S is surcharge, and C is cess.

Tax payments such as TDS, advance tax, and self-assessment tax are considered when determining the remaining amount payable or refund.

$$B = \max(0, T - P) \tag{4}$$

$$F = \max(0, P - T) \tag{5}$$

where P represents recorded tax payments.

The same tax calculation service is used by the calculation interface and return-preparation workflow. This avoids maintaining separate calculation logic in different parts of the application.

V. Return Selection and Preparation

The system uses explicit rules to evaluate the taxpayer profile and identify an appropriate income-tax return form. The decision layer considers income sources, residential status, business or professional income, capital gains, foreign information, and other supported conditions.

The current implementation includes selection logic for ITR-1 through ITR-4 within the supported scope. The result includes the recommended return, eligibility status, reasons, missing information, and unsupported conditions.

For example, a salary-only resident profile within the supported conditions can be considered for ITR-1. A profile containing supported capital gains without business income may require ITR-2, while supported business or professional income can lead to ITR-3 or ITR-4 depending on the applicable conditions.

The important design decision is that return selection is handled by explicit rules rather than by the language model. This allows the reason for a recommendation to be shown to the user.

After the return is selected, the preparation service uses the same deterministic tax result rather than calculating a separate value.

VI. Conversational Onboarding

The conversational interface assists users while they complete their tax profile. It uses the same structured profile as the main form instead of maintaining an independent set of taxpayer information.

The onboarding process can ask about identity, residential status, employment, salary, TDS, other income, deductions, tax payments, documents, and other relevant fields.

Natural-language responses are converted into candidate values and validated before they are stored.

A confirmation step is used before profile mutation. Therefore, an ambiguous response does not automatically overwrite an existing financial value.

VII. Document Intelligence

Documents such as salary slips and Form 16 can contain information that would otherwise have to be entered manually. The application provides a document-processing workflow to extract useful fields from uploaded documents.

The processing sequence is:

1) authenticate and accept the document;
2) extract available text;
3) identify relevant fields;
4) create candidate values with source information;
5) present the extracted information to the user;
6) allow confirmation or correction before updating the profile.

For scanned or textless documents, OCR-related processing is handled according to the available document-processing support. The system does not treat an extracted value as automatically authoritative.

This confirmation step is important because document layouts can vary and an extraction error should not silently affect the final tax calculation.

VIII. Retrieval and AI-Assisted Chat

The conversational layer combines structured taxpayer information, deterministic calculation results, and assessment-year-specific tax knowledge.

The retrieval component provides relevant tax information based on the assessment year and the user's question. This helps the assistant answer questions about tax concepts and applicable rules without relying only on general language-model knowledge.

The general flow is:

User Question → Tax Context
→ Knowledge Retrieval
→ Verified Tax Result
→ AI Explanation

The language model is not treated as the source of numerical tax truth. For calculation-related questions, the structured output of the tax engine remains authoritative.

For example, when a user asks why the calculated tax has a particular value, the assistant can explain the income, deductions, taxable income, regime, rebate, cess, and final liability already produced by the calculation service.

If the AI provider is unavailable or an answer does not satisfy the application's response checks, the system can fall back to a deterministic response.

TABLE I
TECHNOLOGY STACK

Layer	Technology / Role
Frontend	Next.js, TypeScript, Tailwind CSS, Zustand
Backend	FastAPI, Python
Database	PostgreSQL, SQLAlchemy
Tax computation	Python Decimal and deterministic rule modules
Documents	PDF text extraction and document-processing services
PDF output	ReportLab
AI layer	Language-model provider for conversational assistance
Knowledge	Assessment-year-specific retrieval and tax notes
Authentication	JWT-based authenticated API workflow

This design provides the benefits of conversational interaction while keeping high-stakes financial calculations under explicit software rules.

IX. Implementation and Technology Stack

Table I summarizes the main technologies used in the implementation.

The application is organized into frontend, backend, tax-engine, document-processing, return-preparation, and knowledge components. This separation makes it easier to test individual parts and update assessment-year rules without changing the entire application.

X. Evaluation

The evaluation focused mainly on functional behavior and consistency of the implemented workflow rather than claiming general real-world tax accuracy.

The tested areas included:

- tax calculation and regime comparison;
- deductions, rebate, surcharge, and cess behavior;
- return-form selection and rejection conditions;
- document field extraction;
- profile validation and recalculation;
- conversational responses;
- retrieval behavior;
- preparation PDF generation.

A separate salary-based reference comparison was also used to check the calculation path across selected income values under the supported old and new regimes. The comparison produced matching results for the tested cases.

These results demonstrate consistency within the tested scenarios, but they should not be interpreted as proof that every real-world income-tax return will be prepared correctly. The system operates within a defined assessment-year and feature scope, and professional verification remains appropriate before actual filing.

XI. Discussion

The main design decision in the system is the separation between deterministic computation and AI-assisted interaction.

A fully generative tax assistant could provide flexible conversations, but allowing the language model to calculate tax or independently decide return eligibility would make the system harder to verify. Keeping those operations in explicit software rules makes the results reproducible and easier to test.

The retrieval layer adds another useful separation. Tax-related information can be maintained according to the relevant assessment year, while the tax engine continues to perform the numerical calculation. Consequently, changing explanatory knowledge does not require changing the calculation logic.

Document processing follows a similar principle. Extracted information is treated as a candidate and requires user confirmation before it becomes part of the structured tax profile.

Overall, the approach provides a balance between automation and user control. The system can reduce repetitive entry and explain tax-related information without giving the AI model direct authority over the final financial calculation.

XII. Limitations and Future Work

The current implementation is a preparation and assistance system rather than a complete government e-filing platform. Direct submission, e-verification, acknowledgement handling, and filing-status integration are outside the current scope.

Although the return-selection layer covers ITR-1 to ITR-4 within the implemented rules, some advanced schedules and complex taxpayer scenarios remain limited. Additional assessment-year rules and form-specific coverage would be required for broader filing support.

Document extraction also depends on document layout and extraction quality. A larger benchmark containing varied real-world documents would be useful for measuring extraction performance more systematically.

The retrieval component can be extended with broader authoritative tax sources, stronger provenance information, and improved retrieval techniques. Future evaluation can also compare assistant responses using larger controlled datasets and measure user completion time and error rates.

Government e-filing integration would require additional security, authentication, official interfaces, and compliance controls before it could be considered for production use.

XIII. Conclusion

This paper presented a hybrid framework for personalized income tax preparation that combines deterministic tax computation, rule-based return selection, document-assisted information extraction, assessment-year knowledge retrieval, and conversational AI.

The central design principle is to use AI where natural-language interaction is useful while keeping numerical tax decisions and return eligibility within explicit and testable software rules. The resulting system provides a structured workflow in which users can enter or extract tax information, review their profile, understand their tax calculation, identify an appropriate return form, and prepare a return document within the supported scope.

The architecture also provides a foundation for extending assessment-year rules, improving document intelligence, expanding return schedules, and strengthening knowledge-grounded conversational assistance.

ACKNOWLEDGMENT

The authors used generative AI assistance during manuscript drafting, organization, and language editing. The authors are responsible for verifying the technical content, references, experimental results, and final manuscript.

REFERENCES

[1] S. B. J., A. V. K., V. Daimane, C. Shetty, R. Patil, and S. Sethi, "TaxCraft: Machine Learning-Powered Automated Tax Assistant for Personalized Financial Empowerment," in *Proc. 3rd IEEE Int. Conf. Knowledge Engineering and Communication Systems (ICKECS)*, 2025, doi: 10.1109/ICKECS65700.2025.11035348.

[2] R. Kumar, P. Devi, M. K. Mishra, and S. S. Chauhan, "Artificial Intelligence in Goods and Services Network (GSTN) Application: Scope of Implementation," in *Proc. 1st Int. Conf. Advances in Computer Science, Electrical, Electronics, and Communication Technologies (CE2CT)*, 2025, doi: 10.1109/CE2CT64011.2025.10941311.

[3] R. Kumar, R. K. Malhotra, S. Pandey, A. Gehlot, I. Gautam, and S. Chamola, "Role of Artificial Intelligence in Input Tax Credit Reconciliation," in *Proc. 3rd Int. Conf. Pervasive Computing and Social Networking (ICPCSN)*, 2023, doi: 10.1109/ICPCSN58827.2023.00086.

[4] P. Bharati, V. D. Hanchate, M. D, S. Gani, and S. L. Shakkeera, "The Secure Framework to Develop Income Tax Fraud Detection using AI-ML Techniques," *Journal of Emerging Technologies and Innovative Research*, vol. 11, no. 1, pp. b1–b8, Jan. 2024.

[5] N. Naveen Kumar, R. Sridhar, U. Uday Prasanna, and G. Priyanka, "Tax Management in the Digital Age: A TAB Algorithm-based Approach to Accurate Tax Prediction and Planning," in *Proc. Int. Conf. Inventive Computation Technologies (ICICT)*, 2023, pp. 909–913.

[6] Y. Tan, B. Wu, J. Cao, and B. Jiang, "LLaMA-UTP: Knowledge-Guided Expert Mixture for Analyzing Uncertain Tax Positions," *IEEE Access*, vol. 13, pp. 90637–90648, 2025, doi: 10.1109/ACCESS.2025.3571502.

[7] Income Tax Department, Government of India, "Salaried Indi-viduals for AY 2026-27: Returns and Forms Applicable," 2026. [Online]. Available: https://www.incometax.gov.in/iec/foportal/help/ individual/return-applicable-1

[8] Income Tax Department, Government of India, "File ITR-1 (Sahaj) Online User Manual," 2026. [Online]. Available: https://www.incometax.gov.in/iec/foportal/newdownloads/itr/itr-UM

[9] Income Tax Department, Government of India, "Downloads: Income Tax Returns," 2026. [Online]. Available: https://www.incometax.gov.in/iec/foportal/downloads/income-tax-returns

[10] D. S., "TaxWise: AI-Powered Tax Filing Application," GitHub repository, 2026. [Online]. Available: https://github.com/Diganths2005/aaa