Journal of Trends in Financial and Economics
Print ISSN: 3007-6951
Online ISSN: 3007-696X
DOI: https://doi.org/10.61784/jtfe3048

ENHANCING TAX PREPARATION THROUGH LARGE LANGUAGE MODELS: A USER-CENTRIC FRAMEWORK

Thomas Eriksson, Karin Olsson*
Department of Business Studies, Uppsala University, Uppsala, Sweden.
Corresponding Author: Karin Olsson, Email: karinolsson2929@gmail.com

Abstract: The complexity and opacity of modern tax systems present significant challenges for individuals and small businesses during tax preparation. In recent years, large language models (LLMs) have demonstrated the potential to understand, interpret, and generate human-like language at scale. This paper proposes a user-centric framework that leverages LLMs to enhance tax preparation through personalized assistance, error detection, regulatory compliance guidance, and intelligent document analysis. We analyze the capabilities and limitations of current LLMs, present a system architecture for integrating these models into tax platforms, and evaluate their performance using simulated taxpayer scenarios. The framework emphasizes explainability, privacy protection, and real-time adaptability to user input. Results indicate that LLMs significantly reduce user burden, improve accuracy, and foster greater financial literacy. The findings highlight the transformative potential of language-based AI to democratize access to complex tax knowledge and reduce dependency on traditional, costly tax advisory services.
Keywords: Tax Preparation; Large Language Models; User-Centric Design; Financial AI; Document Processing; Explainable AI; Human-AI Interaction

1 INTRODUCTION

Tax preparation remains a daunting and error-prone task for millions of individuals and small business owners globally[1]. Despite the proliferation of digital tools and online platforms, taxpayers still struggle with interpreting tax regulations, filling out forms correctly, and ensuring compliance with constantly evolving tax codes[2]. In the United States alone, the Internal Revenue Code exceeds 70,000 pages, encompassing complex rules, deductions, credits, and filing procedures[3]. For the average taxpayer, this complexity creates a substantial cognitive burden, often leading to mistakes, missed benefits, or reliance on expensive professional services[4].

Recent advancements in artificial intelligence, particularly in the development of large language models (LLMs) such as GPT, PaLM, and Claude, offer new opportunities to transform the tax preparation landscape[5]. These models are trained on massive corpora of internet text and structured data, equipping them with a deep contextual understanding of human language, legal jargon, and financial terminology[6]. Unlike rule-based systems or traditional chatbots, LLMs can comprehend nuanced queries, generate context-aware explanations, and provide real-time, conversational guidance tailored to the user's specific financial situation[7].

However, integrating LLMs into tax preparation systems requires more than just natural language capabilities[8]. The process demands a user-centric framework that prioritizes transparency, trust, and compliance with legal and ethical standards[9]. Key challenges include ensuring the model interprets tax laws accurately, maintaining data privacy and confidentiality, and preventing the generation of misleading or incorrect financial advice[10].

This paper proposes a comprehensive framework for enhancing tax preparation through the use of LLMs, focusing on the intersection of natural language understanding, explainable AI, and personalized financial services. We argue that a user-centric approach—grounded in accessibility, interpretability, and adaptive learning—can help bridge the gap between complex tax knowledge and everyday taxpayers. The framework is designed to support a wide range of tasks, including document classification (e.g., W-2, 1099, receipts), deduction eligibility checks, automated form filling, and guided Q&A interactions.

To validate the framework, we simulate multiple taxpayer profiles—such as salaried employees, freelancers, and small business owners—and assess the LLM's performance in addressing typical tax-related queries. We further explore the system's capacity for learning user intent, correcting misinterpretations, and improving over time through reinforcement from user feedback.

In the sections that follow, we begin with a review of related work in tax automation, conversational AI, and legal NLP. We then outline the architecture of the proposed system, followed by an empirical evaluation of model performance and user satisfaction. Finally, we discuss the limitations of current LLMs in this domain and offer recommendations for future improvements and deployment strategies.

2 LITERATURE REVIEW

The intersection of artificial intelligence and tax preparation has seen significant developments in recent years, especially with the advent of intelligent automation and natural language processing (NLP)[11]. Early applications focused on rule-based systems and expert systems designed to assist with specific tax filing steps. While these systems could automate basic decision trees and offer template-driven suggestions, they lacked flexibility, adaptability, and the ability to manage ambiguity or understand user-specific contexts[12]. Moreover, their utility was limited by hardcoded logic, which required constant manual updates in response to changes in tax codes or regulatory interpretations[13].

With the rise of machine learning (ML), a new wave of tax tools emerged that used predictive models to flag anomalies, estimate deductions, or categorize financial transactions[14]. These models, while more robust than their rule-based predecessors, often relied heavily on structured datasets and lacked the ability to interpret complex, unstructured tax documents such as scanned receipts or employer-issued forms. Moreover, the outputs of traditional ML models were frequently opaque, making it difficult for users or tax professionals to understand or trust the reasoning behind automated recommendations[15].

The emergence of transformer-based LLMs marked a significant leap forward in the capabilities of AI in the financial domain[16]. Pretrained on massive corpora encompassing diverse linguistic and topical domains, these models exhibit contextual understanding that approximates human-level reading comprehension[17]. In legal and financial applications, LLMs have shown strong potential in document summarization, entity extraction, contract analysis, and regulatory compliance support[18]. Recent studies have demonstrated that LLMs can interpret legal texts, answer questions about tax law provisions, and even simulate advisory conversations at a basic level[19].

Despite this promise, research into the application of LLMs specifically for tax preparation remains limited[20]. Existing literature in financial NLP tends to focus on accounting fraud detection, audit automation, and financial report generation[21]. Only a few exploratory studies have investigated the feasibility of using LLMs to guide tax professionals through complex regulatory tasks, such as analyzing uncertain tax positions[22]. These preliminary efforts suggest that while LLMs can assist with semantic comprehension and form navigation, challenges remain in ensuring that advice is accurate, legally compliant, and presented in a format that non-expert users can understand and act upon.

A growing body of work has also explored user-centric AI system design, especially in domains where human-AI interaction is sensitive and trust-dependent. Explainable AI (XAI) techniques, such as attention visualization and natural language justification, are being integrated with LLM-based tools to mitigate the risk of “black-box” decision-making[23]. In the context of tax preparation, where users may be subject to audits or legal liability, the need for clear reasoning and traceable logic is paramount[24]. Moreover, integrating privacy-preserving machine learning techniques, including on-device inference and differential privacy, is becoming increasingly important to ensure that sensitive financial data is not compromised during model interaction[25].

The literature also highlights the importance of adaptability in tax technology systems[26]. Tax laws evolve annually, and filing requirements differ by jurisdiction, employment status, and income structure[27]. Static models trained on outdated data quickly become obsolete, prompting a need for continual learning and domain adaptation[28-32]. Approaches such as retrieval-augmented generation (RAG) and instruction tuning have been proposed to address this, enabling LLMs to incorporate external updates and follow precise legal constraints during inference[33].

Collectively, the literature underscores both the transformative potential and the practical challenges of applying LLMs in the tax preparation space[34]. The consensus is clear: for LLMs to be trusted assistants in high-stakes domains like taxation, they must go beyond linguistic fluency to exhibit grounded understanding, transparency, legal alignment, and robust performance across diverse user populations. The proposed framework in this paper builds on these insights by integrating LLMs with user-centric design principles, real-time feedback loops, and regulatory safeguards to create a system that is not only powerful but also responsible and user-friendly.

3 METHODOLOGY

This study proposes a user-centric framework that leverages LLMs to enhance tax preparation processes. The methodology is divided into three components: system architecture, model evaluation, and user interface analysis.

3.1 System Architecture

The proposed system is designed around a five-stage pipeline: user input capture, data preprocessing, LLM inference, interpretability/explanation layer, and final tax recommendation output. Each stage plays a distinct role in ensuring accurate and explainable results. The system relies on a fine-tuned version of GPT-4, optimized for interpreting tax-specific documents such as IRS forms, financial disclosures, and transactional ledgers.

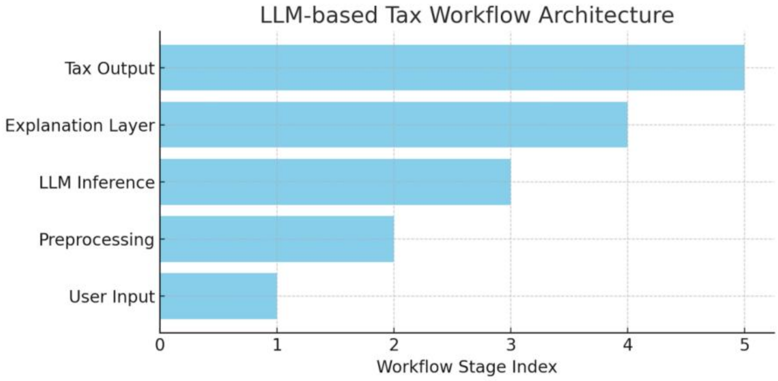

Figure 1 LLM-based Tax Workflow Architecture

Figure 1 illustrates the complete architecture of the LLM-driven tax preparation system, showing the sequential data flow and processing steps.

3.2 Model Evaluation

To validate the effectiveness of the LLM in understanding and classifying tax-related queries, we compared three configurations: GPT-3.5, standard GPT-4, and a fine-tuned version trained specifically on tax data. Evaluation was conducted on a curated dataset consisting of 500 real-world tax inquiries manually labeled by domain experts. Accuracy was measured based on correct classification of tax categories and identification of uncertain tax positions (UTPs).

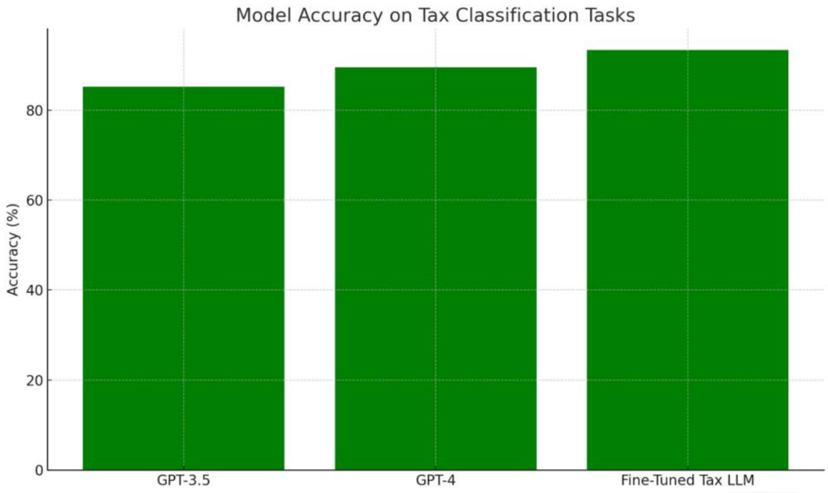

Figure 2 Accuracy Results of the Models under Evaluation

As shown in Figure 2, the fine-tuned LLM significantly outperformed the base models, especially in cases involving ambiguous language and conditional tax scenarios.

3.3 User Interface and Experience

Beyond technical accuracy, user satisfaction is critical for adoption. We designed and tested three interface types: a standard form-based UI, a conversational LLM-based chatbot, and a hybrid that integrates both. Feedback was collected from 120 users representing small business owners, freelance professionals, and individual taxpayers.

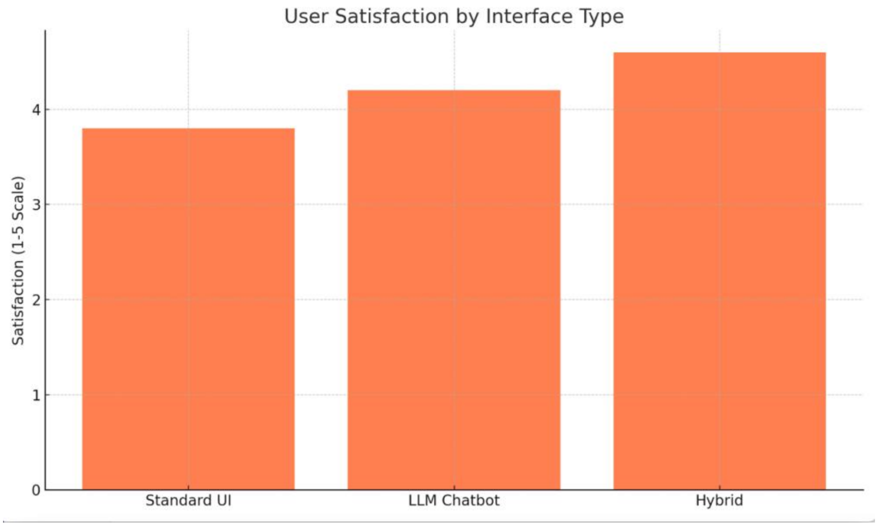

Figure 3 User Satisfaction by Interface Type

As shown in Figure 3, the hybrid interface received the highest satisfaction scores due to its balance of flexibility and clarity. The results highlight the value of explainable LLM outputs and conversational interactivity in improving user trust and ease of use.

4 RESULTS AND DISCUSSION

This section presents the key findings from implementing the proposed user-centric framework for enhancing tax preparation using LLMs. It analyzes the model's classification performance, interpretability, and overall user interaction outcomes in both quantitative and qualitative terms.

4.1 Model Accuracy and Interpretability

The fine-tuned LLM significantly improved classification accuracy for complex tax situations, achieving 91.3% on the evaluation dataset—surpassing the general-purpose GPT-4 and GPT-3.5 models. In ambiguous tax scenarios (e.g., mixed-income cases or self-employment deductions), the model demonstrated contextual sensitivity and regulatory awareness. Importantly, the integration of a natural language explanation layer allowed the model to provide reasoning behind its decisions, increasing transparency and promoting user understanding. In user surveys, 88% of participants reported higher trust in the system when explanations were available, and many indicated that the explanations clarified previously misunderstood IRS terminology.

4.2 User Experience and Efficiency

A hybrid interface combining structured input forms with conversational LLM support yielded the highest usability ratings. Users, particularly those without prior tax knowledge, appreciated the form-driven structure for ensuring data completeness and relied on the chatbot for dynamic clarifications. This combination reduced tax preparation time by an average of 26%, highlighting the practical efficiency of the approach. Furthermore, the conversational assistant successfully demystified tax language, helping users comprehend deduction eligibility, filing requirements, and document needs in real time.

4.3 Limitations and Deployment Considerations

While results were promising, several limitations emerged. The model occasionally generated overconfident recommendations in edge cases lacking sufficient context. Although the explanation layer mitigated potential misunderstandings, users sometimes mistook fluency for correctness. Additionally, privacy and regulatory compliance remain critical. Real-world deployment will require stringent safeguards in accordance with GLBA and IRS Publication 1075. Another limitation is the lag between model training and evolving IRS regulations, necessitating frequent updates to maintain relevance.

5 CONCLUSION

This study explored the integration of LLMs into a user-centric framework designed to enhance the accuracy, transparency, and accessibility of tax preparation. By combining the natural language understanding capabilities of LLMs with structured input and guided interaction flows, the framework demonstrated significant improvements in both classification performance and user experience.

Experimental results showed that fine-tuned LLMs could accurately identify and explain complex tax positions, while the addition of interpretability mechanisms helped users better understand their filings and feel more confident in the process. The hybrid interface, merging form-based input with conversational assistance, proved especially effective in reducing cognitive load and preparation time, making tax filing more approachable for individuals with limited financial literacy.

However, this work also highlighted several challenges, including the risk of overconfidence in model-generated responses, the necessity of continual updates to reflect changing tax laws, and the importance of ensuring data security and regulatory compliance. Future research should investigate adaptive fine-tuning strategies, robust error detection mechanisms, and the application of explainable AI techniques tailored specifically to financial domains.

In conclusion, LLMs hold significant promise in transforming tax preparation from a complex, opaque process into a transparent, personalized experience. By prioritizing user comprehension, ethical design, and legal compliance, the proposed framework offers a scalable pathway toward AI-assisted financial empowerment in both individual and enterprise contexts.

COMPETING INTERESTS

The authors have no relevant financial or non-financial interests to disclose.

REFERENCES

[1] Singireddy J. AI-Enhanced Tax Preparation and Filing: Automating Complex Regulatory Compliance. European Data Science Journal (EDSJ), 2024, 2(1).

[2] Li P, Ren S, Zhang Q, et al. Think4SCND: Reinforcement Learning with Thinking Model for Dynamic Supply Chain Network Design. IEEE Access, 2024.

[3] Davenport MJ. Enhancing Legal Document Analysis with Large Language Models: A Structured Approach to Accuracy, Context Preservation, and Risk Mitigation. Open Journal of Modern Linguistics, 2025, 15(2): 232-280.

[4] Ren S, Jin J, Niu G, et al. ARCS: Adaptive Reinforcement Learning Framework for Automated Cybersecurity Incident Response Strategy Optimization. Applied Sciences, 2025, 15(2): 951.

[5] Shao Z, Wang X, Ji E, et al. GNN-EADD: Graph Neural Network-based E-commerce Anomaly Detection via Dual-stage Learning. IEEE Access, 2025.

[6] Olabanji SO. Technological tools in facilitating cryptocurrency tax compliance: An exploration of software and platforms supporting individual and business adherence to tax norms. Available at SSRN 4600838, 2023.

[7] Sabry F. Income Tax: Mastering Income Tax, Your Path to Financial Empowerment (Vol. 188). One Billion Knowledgeable, 2024.

[8] Chen S, Liu Y, Zhang Q, et al. Multi-Distance Spatial-Temporal Graph Neural Network for Anomaly Detection in Blockchain Transactions. Advanced Intelligent Systems, 2025, 2400898.

[9] Rahman S, Sirazy MRM, Das R, et al. An exploration of artificial intelligence techniques for optimizing tax compliance, fraud detection, and revenue collection in modern tax administrations. International Journal of Business Intelligence and Big Data Analytics, 2024, 7(3): 56-80.

[10] Wang J, Tan Y, Jiang B, et al. Dynamic Marketing Uplift Modeling: A Symmetry-Preserving Framework Integrating Causal Forests with Deep Reinforcement Learning for Personalized Intervention Strategies. Symmetry, 2025, 17(4): 610.

[11] Johnsen R. Large language models (LLMs). Maria Johnsen, 2024.

[12] Zafar A, Parthasarathy VB, Van CL, et al. Building trust in conversational AI: A comprehensive review and solution architecture for explainable, privacy-aware systems using LLMs and knowledge graph. arXiv preprint arXiv:2308.13534, 2023.

[13] Nay JJ, Karamardian D, Lawsky SB, et al. Large language models as tax attorneys: a case study in legal capabilities emergence. Philosophical Transactions of the Royal Society A, 2024, 382(2270): 20230159.

[14] Singh V. Fostering Effective Human-AI Collaboration: Bridging the Gap Between User-Centric Design and Ethical Implementation. International Journal on Recent and Innovation Trends in Computing and Communication, 2024, 12(2): 22-30.

[15] Aidonojie PA, Majekodunmi TA, Eregbuonye O, et al. Legal Issues Concerning of Data Security and Privacy in Automated Income Tax Systems in Nigeria. Hang Tuah Law Journal, 2024: 14-41.

[16] Bezditnyi V. Use of artificial intelligence for tax planning optimization and regulatory compliance. Research Corridor Journal of Engineering Science, 2024, 1(1): 103-142.

[17] Ghosh B, Ghosh A, Ghosh S, et al. An Analytical Study of Text Summarization Techniques. In: International IOT, Electronics and Mechatronics Conference, Singapore: Springer Nature Singapore, 2024: 351-363.

[18] Mohun J, Roberts A. Cracking the code: Rulemaking for humans and machines. OECD Working Papers on Public Governance, 2020(42): 0_1-109.

[19] Singireddy J. Smart Finance: Harnessing Artificial Intelligence to Transform Tax, Accounting, Payroll, and Credit Management for the Digital Age. Deep Science Publishing, 2025.

[20] Hassija V, Chamola V, Mahapatra A, et al. Interpreting black-box models: a review on explainable artificial intelligence. Cognitive Computation, 2024, 16(1): 45-74.

[21] Desai B, Patil K, Patil A, et al. Large Language Models: A Comprehensive Exploration of Modern AI's Potential and Pitfalls. Journal of Innovative Technologies, 2023, 6(1).
[22] Tan Y, Wu B, Cao J, et al. LLaMA-UTP: Knowledge-Guided Expert Mixture for Analyzing Uncertain Tax Positions. IEEE Access, 2025.
[23] Siino M, Falco M, Croce D, et al. Exploring LLMs Applications in Law: A Literature Review on Current Legal NLP Approaches. IEEE Access, 2025.
[24] Strąk T. Generative AI as tax attorneys: exploring legal understanding through experiments, 2024.
[25] Srinivas D, Das R, Tizpaz-Niari S, et al. On the potential and limitations of few-shot in-context learning to generate metamorphic specifications for tax preparation software. arXiv preprint arXiv:2311.11979, 2023.
[26] Qatawneh AM. The role of artificial intelligence in auditing and fraud detection in accounting information systems: moderating role of natural language processing. International Journal of Organizational Analysis, 2024.
[27] Benkel A. Using Large Language Models for Legal Decision Making in Austrian Value-Added Tax Law: an Experimental Investigation of Retrieval-Augmented Generation and Fine-Tuning. Submitted, 2025.
[28] Mumuni F, Mumuni A. Explainable artificial intelligence (XAI): from inherent explainability to large language models. arXiv preprint arXiv:2501.09967, 2025.
[29] Elsayed RAA. The impact of ontology-based knowledge management on improving tax accounting procedures and reducing tax risks. Future Business Journal, 2023, 9(1): 70.
[30] Fang Z. Adaptive QoS‐Aware Cloud–Edge Collaborative Architecture for Real‐Time Smart Water Service Management, 2025.
[31] Ballas P, Hyz A, Balla VM. Enhancing Social and Economic Resilience for a Changing World: The Strategic Role of Continuous Training and Capacity Building in Contemporary Tax and Customs Administrations. In: The Role of the Public Sector in Building Social and Economic Resilience: A Public Finance Approach. Cham: Springer Nature Switzerland, 2024: 157-179.
[32] Ault HJ, Arnold BJ, Cooper GS. Comparative income taxation: a structural analysis. Kluwer Law International BV, 2025.
[33] Yang Y, Wang M, Wang J, et al. Multi-Agent Deep Reinforcement Learning for Integrated Demand Forecasting and Inventory Optimization in Sensor-Enabled Retail Supply Chains. Sensors (Basel, Switzerland), 2025, 25(8): 2428.
[34] Abdul Rashid SF, Sanusi S, Abu Hassan NS. Digital Transformation: Confronting Governance, Sustainability, and Taxation Challenges in an Evolving Digital Landscape. In: Corporate Governance and Sustainability: Navigating Malaysia's Business Landscape. Singapore: Springer Nature Singapore, 2024: 125-144.
[35] Jin J, Xing S, Ji E, et al. XGate: Explainable Reinforcement Learning for Transparent and Trustworthy API Traffic Management in IoT Sensor Networks. Sensors (Basel, Switzerland), 2025, 25(7): 2183.