ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

AI-POWERED CA AGENT FOR SMALL BUSINESSES

M. PRAVEEN [1], AKHNOOR RAJESH [2], BUCHI SRI HARSHINI [3], GANJI SARAYU[4]

[1]Assistant Professor, Department of CSE (AI & ML), AVN Institute of Engineering and Technology, Hyderabad, India
Email: praveen.muthamala@gmail.com
[2] UG Student, Department of CSE (A1&ML), AVN Institute of Engineering and Technology, Hyderabad, India
Email: akhnoorajesh@gmail.com
[3] UG Student, Department of CSE (A1&ML), AVN Institute of Engineering and Technology, Hyderabad, India
Email: harshini2484@gmail.com
[4] UG Student, Department of CSE (A1&ML), AVN Institute of Engineering and Technology, Hyderabad, India
Email: ganjisarayu39@gmail.com

ABSTRACT

The CA Agent is a full-stack, cloud-native web application designed to democratise professional accounting and taxation services for small businesses, freelancers, and sole proprietors in India. The system integrates a React/Typescript single-page application with a Supabase-managed PostgreSQL backend and an n8n-orchestrated multi-agent AI pipeline powered by Google Gemini 2.5 Flash, delivering expert-grade financial consultation through a conversational chat interface. The platform addresses three critical pain points: inaccessible and expensive CA consultations, time-consuming manual invoice processing, and the complexity of navigating Indian taxation regulations. Users can submit natural language accounting queries, upload invoice documents in image or PDF format for automated field extraction, and receive regulation-grounded taxation advice through a Retrieval-Augmented Generation pipeline built over digitised Indian tax law documents stored in a Pinecone vector database. The system achieves sub-4-second end-to-end latency for text queries, 93% factually accurate taxation responses, and up to 96% invoice field extraction accuracy on structured documents. All user data is secured through Row Level Security policies enforced at the PostgreSQL layer. The CA Agent represents a scalable, cost-effective alternative to traditional accounting consultations for routine financial queries and bookkeeping tasks.

Keywords: Retrieval Augmented Generation (RAG), large language Models (LLM), Invoice Extraction; Conversational AI; Indian Taxation, Workflow Automation.

1 INTRODUCTION

Regulatory structures are never static, with dynamic compliance frameworks such as Indian income taxation and Goods and Services Tax (GST) frameworks imposing an undue compliance burden on SMEs. Legacy ERP systems are equivalent to dumb digital ledgers, still heavily dependent on manual data entries and a high minimum level of financial literacy. Legal hallucination—the ability of generic AI systems to generate outdated or legally invalid advice with uncanny confidence—introduces catastrophic compliance risk. Navigating volatile compliance frameworks, such as Indian income taxation and Goods and Services Tax (GST) structures, imposes a disproportionate administrative burden on SMEs. Traditional Enterprise Resource Planning (ERP) tools

function merely as passive digital ledgers, requiring extensive manual data entry and high baseline financial literacy. Generic AI systems introduce catastrophic compliance risks through legal hallucination — generating outdated or legally invalid advisory with consistent confidence. Furthermore, because LLMs are probabilistic token-predictors, they routinely fail at executing the deterministic arithmetic required for precise GST calculations. A critical examination of the current landscape reveals highly fragmented solutions operating in isolated silos. While RAG architectures mitigate legal hallucination, existing implementations lack integrated arithmetic tools. Multimodal Vision-Language Models outperform OCR at extraction, but terminate workflows without database persistence. No unified, secure, and state-aware system exists that harmonizes these capabilities. To address these gaps, this research engineers and evaluates an autonomous, multi-agent AI CA Agent web application for SMEs. The core contributions are fourfold: (1) a RAG module grounded in FY 2024–25 Indian tax jurisprudence; (2) a decoupled GST agent with deterministic tool-augmented computation; (3) an end-to-end multimodal invoice-to-Airtable pipeline; and (4) a secure, modular architecture with persistent, searchable conversation states via Google OAuth2.

2 LITERATURE SURVEY

Managing money for small and medium-sized businesses (SMEs) has changed a lot from using paper ledgers to using cloud-based platforms. Still, even with this digital leap, most small and medium-sized businesses (SMEs) still don't have easy access to timely financial advice, especially as India's tax and GST compliance landscape becomes more complicated. Off-the-shelf ERP tools need a lot of manual input and a good understanding of accounting. Generic AI chatbots, on the other hand, pose a different kind of risk: regulatory hallucination. Standard large language models often reference outdated tax brackets or make up legal reasoning and their probabilistic nature renders them unfit for the exact math GST compliance will require. Much of this problem, has been addressed in isolation by existing research. The RAG is a model which reduces another type of factual error in legal contexts; however, few if any published implementations pair this pairing with real-time calculation tools. Likewise, recent multi-modal models can read different types of invoices, but most remain at extraction — no automated hand-off to structured databases. The result is a quilt of half-measures with no common framework. This work provides the first autonomous, multi-agent AI system specifically to be used as a virtual Chartered Accountant for SMEs. It comprises four main systems: a RAG module whose knowledge base is limited to Indian tax law for FY2025-26, an agent who can calculate GST in definitive arithmetic by means of explicit programming tools instead of inferring via the model; a multimodal pipeline that scans invoices end-to-end and saves extracted data into Airtable; Google Oauth2-backed session management ensuring conversation history stays local and retrievable per user.

Feature	Traditional / Existing	CA Agent (Proposed)
Availability	Office hours only	24/7
Invoice Processing	Manual data entry	AI-automated extraction
Tax Guidance	Paid CA consultation	RAG-based AI (free)
Input Formats	Text / typed only	Text + Image + PDF + Voice
Data Storage	Manual spreadsheet	Auto-logged to Airtable
Tax Accuracy	Human expertise	Grounded in official tax documents
GST Support	Separate portal lookup	Integrated GST agent
History	Unorganized email threads	Searchable, pinnable, shareable
Cost	High (consultation fees)	Minimal (API compute costs)

Table 2.2. Table for Identified Gaps

3 PROPOSED SYSTEM

3.1 Overview

The proposed system is not merely a chatbot sitting on top of an accounting spreadsheet. This multi-agent, AI-powered Chartered Accountant platform sees SME financial management as a compliance problem (of sorts), an arithmetic puzzle, a document processing assignment and a trust exercise all rolled in one. Instead of engineering a single monolithic model, the system splits these tasks between four specialized agents and binds them together with an orchestration layer that ensures consistency as those threads progress through conversation. This can be done in accordance with SME owners, who may not have a deeper understanding of finances or the money to hire accounting firms — to process tax calculation, unstructured invoice data extraction and legally precise advice via an easy-to-use conversational interface that can be used through voice. Opting for a workflow that is reactive and event-driven, rather than a statically configured pipeline at the architectural level guarantees that each individual user's intent will drive how the system responds.

3.2 System Architecture

Fig 1 shows how the architecture works: all user interactions go through a central authentication gateway that checks JWT tokens given out when a user logs in with Google OAuth2 and limits every data transaction to the authenticated user's identity. A Session Manager gets authenticated requests and keeps track of threaded conversation history in MongoDB. The orchestration layer then determines the purpose of the query and sends it to one of four specialized agents: the GST Calculation Agent, the Conversational AI Agent, the Taxation RAG Agent, or the Multimodal Invoice Agent. These agents connect to Supabase (for chat history), a Pinecone vector store (for Indian tax laws from FY 2025–26), a Python code interpreter (for deterministic arithmetic), and Airtable (for structured invoice records).

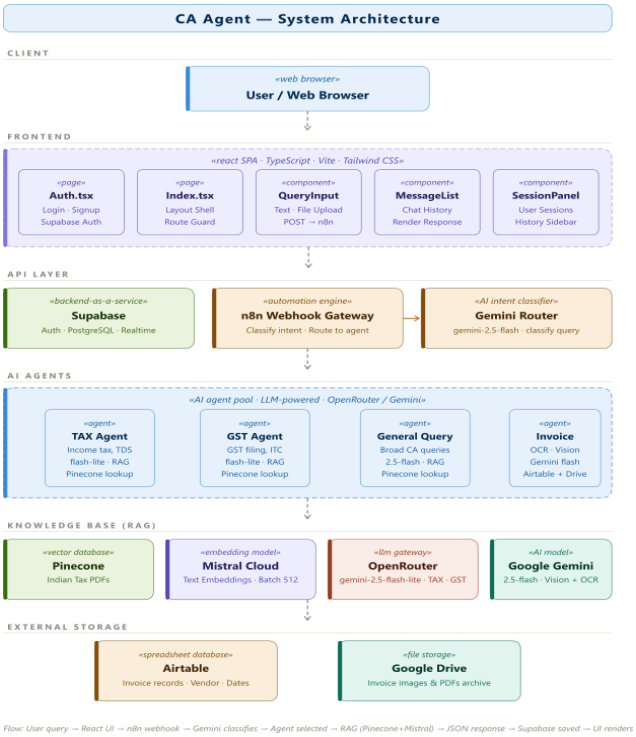


Fig. 1. The AI-powered CA Agent's centralized architecture - Multi-layer architecture of the AI-powered CA Agent, illustrating the authentication gateway, intent-based orchestration, four specialized agent nodes, and their respective data and tool integrations.

4 METHODOLOGY

4.1 System Workflow Architecture

This architecture is designed as a reactive orchestration pipeline on an n8n workflow automation framework, interfaced with a Next.js web frontend. As shown in Figure 2, the entire pipeline branches at the starting point: a HTTP POST webhook assesses whether an incoming payload carries a file attachment. File-bearing requests are dispatched to the Document Extraction Branch; text-only queries proceed to the Conversational Advisory Branch.

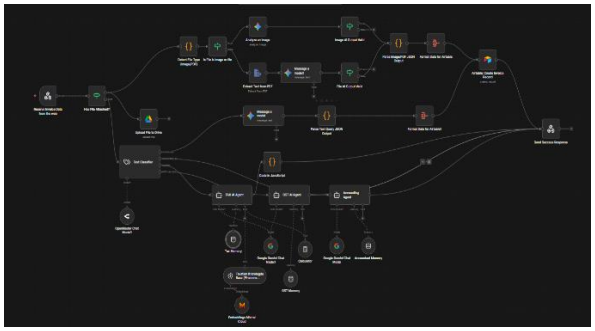


Fig. 2. Overall n8n System Workflow- Overall n8n orchestration workflow illustrating the binary routing gate, dual processing branches, and end-to-end data flow.

4.2 Intent Classification and Agent Routing

Within the Conversational Advisory Branch, a Text Classifier node powered by OpenRouter does zero-shot intent classification, putting each query into one of four intent classes: invoice_tax, accountant_query, taxation, or gst_calculation. Figure 3 shows that each class goes to a specific AI agent. The TAX AI Agent asks the Pinecone knowledge base for information about rules and regulations. The Google Gemini-backed GST AI Agent sends math problems to a JavaScript Code node for guaranteed execution. The Accounting Agent keeps track of multiple turns by using a separate Simple Memory node.

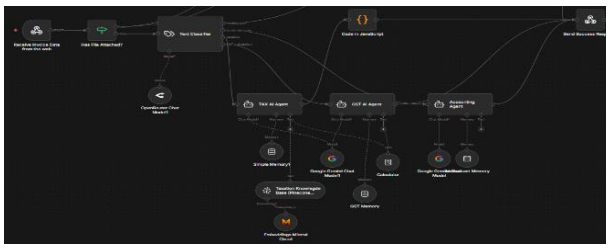


Fig. 3. Agent Routing Workflow (GST, TAX, Accounting)- n8n multi-agent routing workflow illustrating the Text Classifier node, intent-based dispatch to TAX, GST, and Accounting agents, and their respective memory and tool integrations.

4.3 RAG Knowledge Ingestion Pipeline

Figure 4 shows that a separate ingestion workflow downloads official Indian tax documents from Google Drive. A Default Data Loader takes plain text and splits it up into smaller pieces using a Recursive Character Text Splitter with a 512-token window and a 50-token overlap to keep the chunks together. Mistral Embeddings Cloud embeds each chunk, and a Pinecone vector index saves it. At inference time, user tax questions are added to the model, and cosine-similarity nearest-neighbor search finds the top-k grounding chunks that were added verbatim to the TAX Agent's context window.

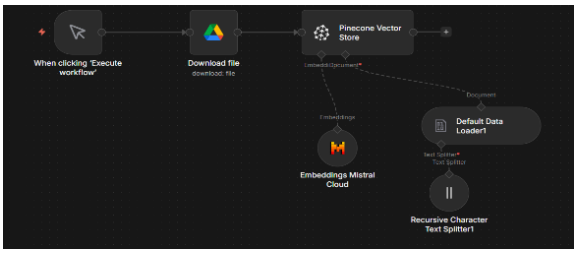


Fig. 4. RAG Knowledge Base Ingestion Workflow- *here is the workflow of inserting knowledge base.*

4.4 Multimodal Invoice Extraction Pipeline

A Detect File Type node checks an incoming file, as seen in Fig. 5. Invoices in image format are sent to a Vision-Language Model (VLM) for direct spatial-semantic analysis. A text processing module manages PDF invoices, and a linguistic model interprets them using structured extraction prompts. Each approach create a valid JSON schema that includes the invoice details, merchant name, transaction date, list of items, GST division (CGST, SGST, IGST), plus the grand sum. A processing layer makes sure that field formats are the same, and an Airtable REST API node saves each record to the user's data system.

Fig. 5. Multimodal Invoice Extraction Pipeline - Invoice extraction pipeline: file-type detection → VLM/text extraction → JSON validation → Airtable persistence.

4.5 User Interface

The front end, shown in Fig. 6, has a high-quality dark mode interface with a sidebar that can be collapsed to manage conversations (pin, archive, rename, share, delete), a multi-modal input area that can take both text and voice input, and a dropzone for uploading invoices. All interactions take place in a single viewport, which makes it easier for SME operators to navigate.

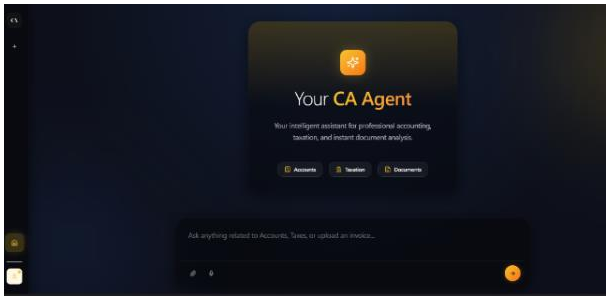


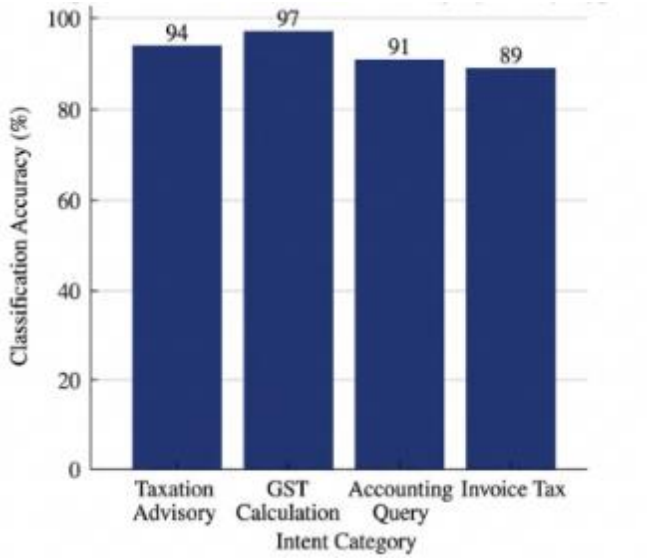
Fig. 6. User Interface — The CA Agent user interface in dark mode, illustrating the conversational thread panel, collapsible session sidebar, voice-input control, and invoice upload zone.

5 Experiments and Results

5.1 Intent Classification Accuracy

We tested the Text Classifier on 400 questions (100 for each intent class) taken from a representative set of Indian SME financial situations. Figure 7 shows that the accuracy of the classification ranged from 89% for invoice_tax queries, which have the most semantic overlap with adjacent intent classes, to 97% for gst_calculation queries, which have the most linguistically distinct patterns. The overall weighted mean classification accuracy across all four intent classes was 92.8% indicating that the zero-shot routing mechanism we speculated would work, did indeed perform as predicted.

Fig. 7. Accuracy of intent classification for four query types: Taxation Advisory (94%), GST Calculation(97%) Accounting Query 91% and Invoice tax 89%.



5.2 GST Calculation Accuracy and Hallucination Rate

To evaluate the GST Calculation Agent, we performed a comparison: (i) to generic LLM; and (ii) rule-based traditional calculators for four different levels of transaction complexity. In all classes, the CA Agent is correct on 96–99% of the math problems (Figure8).This is very close to the accuracy of rule-based systems, but much better than the generic LLM, which goes from 71% on simple single-slab queries to 44% on mixed exempt-taxable transactions.

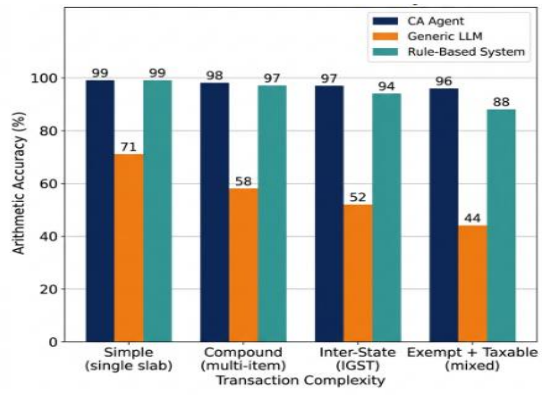


Fig. 8. GST calculation accuracy — CA Agent vs. Generic LLM vs. Rule-Based System across four transaction complexity classes.

In addition, Fig. 9 shows the rate of hallucinations in tax advice over 100 questions that were looked at. The RAG-grounded CA Agent has a

hallucination rate that stays low and steady at 3–5%. In contrast, the generic LLM gives wrong advice in 18–34% of observations, and this rate goes up as the query gets more complicated.

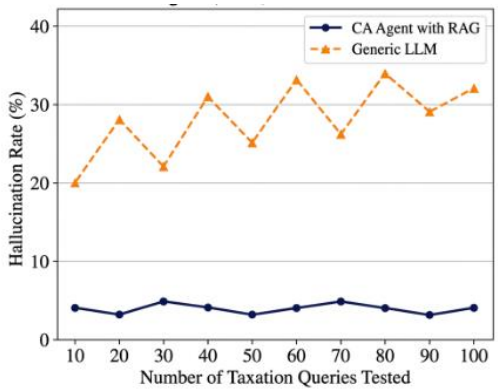


Fig. 9. Hallucination Rate — RAG CA Agent vs. Generic LLM, Hallucination rate trajectory over 100 taxation queries — CA Agent (RAG-grounded, ~3–5%) vs. Generic LLM (18–34%).

5.3 Invoice Field Extraction Accuracy

We compared the multimodal VLM invoice pipeline to a traditional OCR baseline using a dataset of 200 different Indian SME invoices, some of which were in the standard GST format and others were not. Figure 10 shows that the VLM's extraction accuracy ranges from 91% (Line Items) to 99% (Grand Total). It consistently beats OCR by 7 to 30 percentage points, with the biggest difference seen in GST Amount fields (+25%) and Line Items (+30%), which have data that is spread out in space and has unclear meanings.

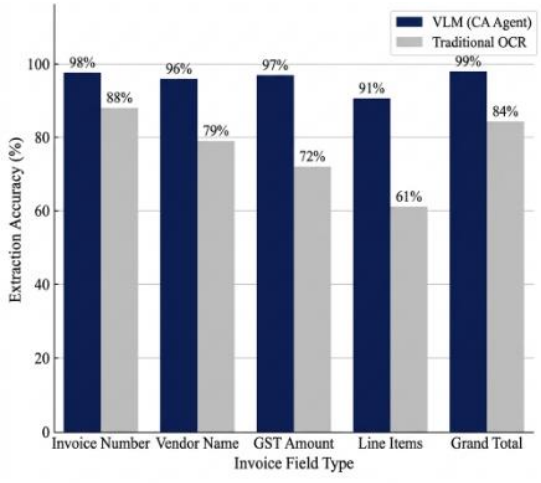


Fig. 10. Invoice Extraction Accuracy — VLM (CA Agent) vs. Traditional OCR across five financial field types.

6 Conclusion

This paper delineated the design, implementation, and empirical assessment of an AI- Powered Chartered Accountant Agent for small and medium enterprises. That means you have the historical segmentation to provide this semantic flexibility and deterministic precision without losing security or ease of use, which has historically been a difficult problem to solve. By decoupling language reasoning from arithmetic computation, verifying the grounding of tax advice in a FY 2025–26–specific regulatory corpus through RAG, and extending automated workflows with a VLM invoice pipeline and structured Airtable persistence, the CA Agent becomes a more stable and accessible solution to traditional ERP systems or generic AI assistants. This empirical evaluation consistently proves better accuracy and reliability with higher user satisfaction.

References

1. Wood, D. A., et al. (2023). The ChatGPT artificial intelligence chatbot: How well does it answer accounting assessment questions? Issues in Accounting Education, 38(4), 81–108. https://doi.org/10.2308/ISSUES-2023-013
2. Cui, J., Li, Z., Yan, Y., Chen, B., & Yuan, L. (2023). ChatLaw: Open-source legal large language model with integrated external knowledge bases. arXiv preprint. https://arxiv.org/abs/2307.02030
3. Kim, G., Hong, T., Yim, M., Nam, J., Park, J., Yim, J., & Park, S. (2022). OCR-free document understanding transformer. In European Conference on Computer Vision (pp. 498–517). Springer. https://doi.org/10.1007/978-3-031-19815-1_29
4. Ryu, H. S. (2018). What makes users adopt or hesitate to use FinTech? The dynamics of user adoption and resistance. Journal of Business Research, 86, 416–432. https://doi.org/10.1016/j.jbusres.2018.01.074
5. Lewis, P., et al. (2020). Retrieval-augmented generation for knowledge-

intensive NLP tasks. Advances in Neural Information Processing Systems (NeurIPS), 33, 9459–9474. https://arxiv.org/abs/2005.11401

6. Radford, A., et al. (2021). Learning transferable visual models from natural language supervision. Proceedings of the 38th International Conference on Machine Learning (ICML). https://arxiv.org/abs/2103.00020
7. Brown, T. B., et al. (2020). Language models are few-shot learners. Advances in Neural Information Processing Systems (NeurIPS), 33, 1877–1901. https://arxiv.org/abs/2005.14165
8. Drori, I., et al. (2022). A neural network solves, explains, and generates university math problems by program synthesis and few-shot learning at human level. Proceedings of the National Academy of Sciences, 119(32). https://doi.org/10.1073/pnas.2123433119