

TaxIQ

India's Unified Tax Intelligence Agent

From Invoice Photo to ITR — One AI Platform

STARTUP BUSINESS PLAN & TECHNICAL ARCHITECTURE

2025 | Confidential

GST Filing in 5 Min

AI Fraud Detection

Max Your Tax Savings



#76 • FINTECH / GOVTECH / GRAPH AI

Intelligent GST Reconciliation Using Knowledge Graphs



PROBLEM OVERVIEW

Build a Knowledge Graph-based GST reconciliation engine that models GST data as graph entities, enables multi-hop traversal to validate invoice-to-tax-payment chains, classifies mismatches by financial risk, and generates explainable audit trails. India's GST reconciliation is fundamentally a graph traversal problem — not flat table matching. The solution addresses ITC leakage affecting 1.4 crore taxpayers by modeling GSTR-1, GSTR-2B, GSTR-3B, e-Invoice, and e-Way bill data as interconnected graph nodes.

KEY OBJECTIVES

- ✓ Model GST data as graph entities and relationships
- ✓ Enable multi-hop traversal for ITC validation
- ✓ Classify and prioritize mismatches by financial risk
- ✓ Generate explainable audit trails for discrepancies
- ✓ Predict vendor compliance risk using graph patterns

REQUIREMENTS

- Knowledge Graph schema for GST ecosystem (Taxpayers, GSTINs, Invoices, IRNs, Returns)
- Mock data ingestion for GSTR-1, GSTR-2B, Purchase Register, e-Invoice
- Graph traversal reconciliation engine with root-cause classification
- ITC risk dashboard with vendor compliance scores
- Explainable natural-language output for audit support
- Predictive vendor risk scoring using historical graph patterns

DELIVERABLES

- 1 Knowledge Graph schema and GST data model
- 2 Reconciliation engine with mismatch classification
- 3 Interactive ITC risk and vendor compliance dashboard
- 4 Explainable audit trail generator
- 5 Predictive vendor compliance risk model

1. Executive Summary

TaxIQ is an AI-powered, three-in-one tax intelligence platform purpose-built for India's 63 million MSMEs and 300 million salaried taxpayers. It unifies GST invoice filing, fraud detection via knowledge graphs, and personal income tax optimization into a single mobile-first agent accessible via WhatsApp, SMS, and a web dashboard.

India loses over ₹1.5 lakh crore annually to fraudulent GST Input Tax Credit (ITC) claims, while simultaneously 72% of eligible taxpayers leave significant deductions under Sections 80C and 80D unclaimed. Both problems share a root cause: no single platform connects raw financial data — invoices, bank statements, spending patterns — with actionable, personalized tax intelligence.

TaxIQ solves this with three deeply integrated AI agents:

- The GST Filing Agent reads invoice photos using OCR + LLM, extracts structured data, and auto-fills GSTR-1 returns in under 5 minutes
- The Fraud Detection Engine builds a live Knowledge Graph of GSTIN relationships to detect circular ITC fraud, mismatch patterns, and anomalous claim chains before filing
- The Tax Optimizer Agent parses bank statement CSVs, identifies investment gaps under 80C/80D/NPS, compares Old vs. New regime, and generates a personalized savings roadmap

Metric	Value
Target Market Size (SAM)	₹8,400 Cr/year
Primary User Segment	MSMEs + Salaried Professionals
Core Problem Solved	Tax compliance + optimization complexity
Time to File GST (Before TaxIQ)	4–6 hours
Time to File GST (With TaxIQ)	< 5 minutes
Projected Year 1 Revenue	₹1.2 Cr (conservative)
Breakeven Target	Month 18

2. The Problem

2.1 India's Tax Compliance Crisis

India's tax ecosystem is fractured across three layers — each underserved by existing tools:

Layer	Who Suffers	Scale of Problem	Current Solution
GST Invoice Filing	Kirana stores, freelancers, MSMEs	63M registered GST taxpayers	Manual entry in Tally / CA help
ITC Fraud & Reconciliation	GSTN, honest taxpayers, govt	₹1.5L Cr+ lost annually	Post-audit by GSTN officers
Income Tax Optimization	Salaried + self-employed	₹78,000 avg unclaimed deductions	CA consultation (₹5–15K/year)

Input Tax Credit (ITC) claims, prevents tax leakage, and avoids notices or penalties

2.2 The Pain Points in Detail

For Small Business Owners

- GST compliance requires understanding of HSN codes, CGST/SGST/IGST splits, GSTR-1 vs GSTR-3B differences
- Average Kirana store spends ₹3,000–8,000/year hiring a local CA just for GST filing
- Errors in filing lead to notices, penalties, and cascading reconciliation problems
- Language barrier — most software is English-only; shop owners operate in Hindi, Tamil, Telugu, Marathi

For Individual Taxpayers

- 72% of taxpayers do not fully utilize Section 80C limits (₹1.5 lakh)
- 80D (health insurance deduction) is largely unknown among under-35 earners
- Old vs New regime decision is made without proper comparison — most choose wrong
- Tax planning is reactive (March rush) rather than year-round and proactive

For the Tax System

- Circular ITC fraud is systematic: shell companies create fake invoice chains to claim credits
- GSTN detects most fraud post-filing, leading to complex audit cycles
- There is no real-time pre-filing fraud scoring visible to compliant filers

3. The Solution — TaxIQ Platform

3.1 Product Overview

TaxIQ is not a tax filing tool. It is a tax intelligence agent — an always-on AI that sits between raw financial data and the Indian tax system, translating chaos into compliance and savings.

3.2 Layer 1 — GST Invoice Intelligence Agent

The filing agent accepts invoice inputs through multiple channels:

- WhatsApp photo forward (most accessible for Kirana stores)
- Mobile camera upload via PWA
- Scanned PDF from email
- Bulk CSV import for larger businesses

The agent pipeline:

- OCR extracts raw text from invoice using Google Cloud Vision API
- Claude AI parses the raw text into structured JSON: vendor GSTIN, HSN codes, CGST/SGST/IGST amounts, invoice number, date
- Validation layer checks GSTIN format, HSN code validity, tax calculation accuracy
- Auto-populates GSTR-1 JSON payload ready for GSTN submission
- Explains every field in the user's preferred language (Hindi, Tamil, Telugu, Marathi, Kannada)

3.3 Layer 2 — Knowledge Graph Fraud Detection Engine

Every invoice processed by TaxIQ becomes a node in a Neo4j knowledge graph. The graph continuously maps:

- GSTIN nodes: each registered business entity
- Invoice nodes: each document with value, date, tax amounts
- ITC Claim edges: who claimed credit from whom
- Supply edges: who sold to whom

The fraud detection algorithms run on this graph in real-time:

- Cycle Detection: identifies circular ITC chains (A claims from B claims from C claims from A)
- Star Pattern Detection: one GSTIN claiming ITC from suspiciously many small vendors simultaneously
- Velocity Scoring: GSTINs with sudden spikes in invoice volume get flagged

- Mismatch Alerts: buyer's claimed ITC vs seller's declared supply — discrepancy > 5% triggers warning

The output is a Compliance Confidence Score (0–100) shown to every user before they file. This is the first consumer-facing, pre-filing fraud score in India.

3.4 Layer 3 — Personal Tax Optimizer Agent

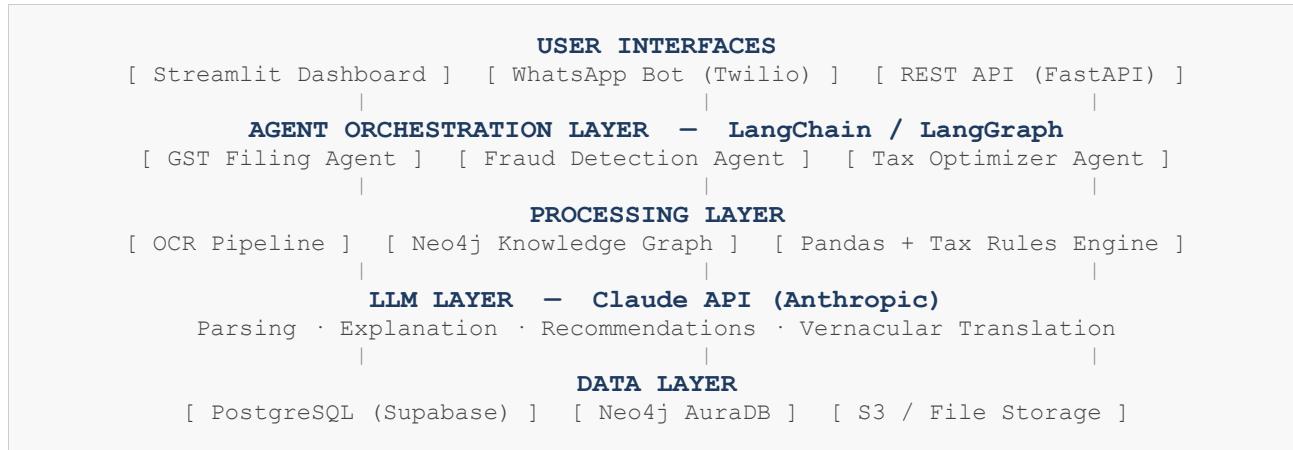
The tax optimizer ingests bank statement CSVs (any format — SBI, HDFC, ICICI, Axis, Kotak) using an intelligent Pandas pipeline that:

- Auto-categorizes transactions: LIC premiums, SIP investments, health insurance, tuition fees, home loan EMIs, medical bills
- Maps each category to the applicable tax section
- Calculates current investment per section vs. statutory limit
- Identifies exact rupee gap and remaining months in financial year
- Suggests specific instruments (ELSS vs PPF vs NPS vs NSC) based on liquidity needs, risk profile, and time horizon
- Runs Old Regime vs New Regime comparison with actual numbers
- Generates a month-by-month investment calendar to spread savings evenly

Unique cross-layer enrichment: business purchases and equipment costs visible in Layer 1's GST data are automatically surfaced as potential deductions in Layer 3 — something no standalone tool can do.

4. Technical Architecture

4.1 System Architecture Overview



4.2 Technology Stack

Component	Technology	Purpose	Cost
LLM / AI Core	Anthropic Claude API	Invoice parsing, tax advice, vernacular explanation	~₹0.25/invoice
OCR	Google Cloud Vision API	Extract text from invoice images	Free up to 1K/mo
Knowledge Graph	Neo4j AuraDB	GSTIN relationship mapping, fraud detection	Free tier available
Database	PostgreSQL on Supabase	Users, invoices, transactions, tax profiles	Free tier available
Backend API	FastAPI (Python)	REST endpoints, agent orchestration	Self-hosted
Agent Framework	LangChain + LangGraph	Multi-agent coordination, tool use	Open source
Data Processing	Python + Pandas	Bank statement parsing, tax calculations	Open source
Frontend	Streamlit	Dashboard for demo and power users	Open source
WhatsApp Bot	Twilio API	Mass-market mobile interface	Pay per message
File Storage	AWS S3 / Cloudflare R2	Invoice image storage	~₹2/GB/month

Auth	Supabase Auth / JWT	User authentication and sessions	Free tier
Deployment	Docker + Railway/Render	Container orchestration	₹2K–5K/month

4.3 Data Flow

Invoice Photo Flow: User uploads photo via WhatsApp/web → Google Vision OCR extracts raw text → Claude API parses structured JSON → Validation layer checks GSTINs and tax math → Stored in PostgreSQL → Graph node created in Neo4j → GSTR-1 payload built → User approves → Submitted to GSTN API

Bank Statement Flow: User uploads CSV → Pandas parser normalizes bank format differences → Transaction classifier maps each row to tax categories → Gap analyzer computes investment shortfalls per section → Claude generates personalized recommendations → PDF report generated → Cross-reference with GST data for business deductions

Fraud Detection Flow: New invoice added to graph → Cypher queries run cycle detection → Velocity and mismatch algorithms score all connected GSTINs → Risk scores updated → Compliance confidence score displayed to user before filing → High-risk patterns flagged for review

4.4 Database Schema (Key Tables)

Table	Key Fields	Purpose
users	id, phone, gstin, pan, annual_income	Core user profile
invoices	vendor_gstin, buyer_gstin, hsn_code, cgst, sgst, igst, confidence_score	Parsed invoice data
gstr1_returns	user_id, filing_period, json_payload, compliance_score, fraud_flags	Filed and draft returns
bank_transactions	user_id, txn_date, amount, category, tax_section	Classified transactions
tax_profiles	invested_80c, invested_80d, invested_nps, regime_choice, estimated_tax	Annual tax planning state

5. Market Analysis

5.1 Total Addressable Market (TAM / SAM / SOM)

Market Level	Definition	Size	Revenue Potential
TAM — Total Addressable Market	All GST-registered businesses + income tax filers in India	95M+ entities	₹42,000 Cr/year
SAM — Serviceable Addressable Market	MSMEs + salaried professionals with digital access	63M businesses + 80M individuals	₹8,400 Cr/year
SOM — Serviceable Obtainable Market	Year 3 target: 0.5% of SAM via digital + partner channels	~715,000 users	₹85 Cr/year

5.2 Market Segmentation

Segment	Size	Willingness to Pay	Primary Pain	Channel
Micro businesses (Kirana, street vendors)	40M+	Low (₹99–299/mo)	GST filing complexity	WhatsApp, local partners
Small businesses (1–20 employees)	20M+	Medium (₹499–999/mo)	Filing + fraud risk	Web, accountant networks
Freelancers / Self-employed	15M+	Medium (₹299–699/mo)	ITR + GST dual compliance	App, social media
Salaried professionals	80M+	Low-medium (₹199–499/mo)	Tax saving optimization	Corporate, fintech apps
Chartered Accountants / Tax Firms	500K+	High (₹2,000–5,000/mo)	Efficiency for multiple clients	B2B direct sales

5.3 Market Trends Supporting TaxIQ

- India's GST collection crossed ₹2 lakh crore/month in 2024 — compliance enforcement is intensifying
- Government's push for e-invoicing mandates is cascading down to smaller business tiers
- WhatsApp has 500M+ active users in India — making it the most accessible enterprise interface
- India's fintech market is expected to reach \$1.5 trillion by 2025 — tax tech is underrepresented
- NPS and 80CCD(1B) awareness is growing rapidly among young professionals aged 25–35
- GST Council's real-time matching proposals signal increasing need for pre-filing fraud checks

6. Competitive Analysis

6.1 Current Landscape

Competitor	What They Do	Strength	Critical Weakness
ClearTax	GST filing + ITR filing platform	Brand recognition, large user base	No AI invoice parsing, no fraud detection, no cross-linking between GST and ITR data
Zoho Books	Accounting software with GST module	Deep accounting features, SMB focus	Expensive, requires manual data entry, no AI recommendations
QuickBooks India	Accounting + GST compliance	Global brand, feature-rich	Not WhatsApp-native, complex UI, not vernacular, no fraud layer
TaxBuddy	ITR filing + tax saving guidance	Simple UI, personal finance angle	No GST features, no invoice parsing, no knowledge graph, no fraud detection
GSTN Portal	Official government filing portal	Free, authoritative	No intelligence, no recommendations, no fraud pre-screening
CAs / Local Accountants	Manual compliance services	Trusted, personalized advice	Expensive (₹5K–20K/year), slow, not scalable, no real-time intelligence

6.2 TaxIQ Differentiation Matrix

Feature	TaxIQ	ClearTax	Zoho Books	TaxBuddy
Invoice photo → GSTR-1 (OCR + AI)	Yes	No	No	No
Knowledge Graph fraud detection	Yes	No	No	No
Pre-filing Compliance Score	Yes	No	No	No
Bank statement → 80C/80D gap analysis	Yes	Partial	No	Partial
Old vs New regime auto-comparator	Yes	Yes	No	Yes
WhatsApp-native interface	Yes	No	No	No
Vernacular language support	Yes	Partial	No	No
GST + ITR cross-data enrichment	Yes	No	No	No
Pricing (entry tier)	₹99/mo	₹2,499/yr	₹4,200/yr	₹999/yr

6.3 Unique Moats

- Knowledge Graph Network Effect: The more users file through TaxIQ, the richer the GSTIN relationship graph, the better fraud detection becomes for all users — a classic data network moat
- WhatsApp-first Architecture: No competitor has built a genuinely WhatsApp-native filing experience; this gives access to 40M+ MSMEs who never use browser-based tools
- Cross-layer Deduction Discovery: The linkage between GST purchase data and income tax deductions is technically novel and invisible to any tool that handles only one side
- Vernacular + Low-Literacy Design: Built from the ground up for non-English, low-digital-literacy users — ClearTax and Zoho have retrofitted this as an afterthought

7. Revenue Model

7.1 Pricing Tiers (B2C)

Plan	Price	Target User	Key Features
Starter (Free)	₹0/month	First-time filers, students	5 invoices/month, basic GSTR-1, 80C summary only
Solo	₹99/month	Freelancers, small traders	50 invoices/month, full GST filing, basic tax optimizer, Hindi support
Business	₹299/month	MSMEs with 5–50 invoices/day	Unlimited invoices, compliance score, full tax optimizer, 5 languages, WhatsApp bot
Professional	₹699/month	Small businesses, self-employed	Everything + ITC fraud alerts, investment calendar, priority support, CA connect

7.2 B2B Revenue Streams

Stream	Model	Price Point	Target Customer
CA/Accountant Pro	Per-CA subscription	₹2,000–5,000/month	500K chartered accountants managing multiple clients
Bank/NBFC Integration	API licensing + revenue share	₹10–50 per loan assessment	Banks using GST data for MSME credit scoring
Enterprise Payroll	Per-employee/month	₹30–80/employee/month	HR teams offering tax optimizer as employee benefit
Government / GSTN Partnership	Annual license	₹1–5 Cr/year	GSTN / state GST departments for fraud analytics
White Label	Platform licensing	Custom pricing	Regional banks, fintech startups, NBFCs

Targeted Customers:

1. GST Suvidha Providers (GSPs)

Companies officially licensed by GSTN to provide GST filing services.

Who they are: Tally Solutions, ClearTax, Zoho Books, BUSY Accounting, Vyapar

Why they buy: They already have millions of filers on their platform. Adding a fraud/mismatch detection layer makes their product 10x stickier and premium-worthy. They would white-label your engine.

Willingness to pay: ₹50L–₹5Cr annually for a B2B API license

2. Chartered Accountant Firms & Tax Consultancies

India has **1.4 lakh registered CA firms** managing GST for lakhs of businesses.

Who they are: Big4 (Deloitte, EY, KPMG, PwC) tax practices, mid-size CA firms (50–500 clients), solo CAs managing 100+ GST filings monthly

Why they buy: CAs are personally liable if a client's ITC claim is later rejected. A tool that catches mismatches *before* filing protects them legally and saves hours of manual reconciliation in GSTR-2A/2B matching.

Willingness to pay: ₹2,000–₹15,000/month per firm

3. Large Enterprises with Complex Supply Chains

Companies with hundreds of vendors where ITC reconciliation is a monthly nightmare.

Who they are: FMCG companies (HUL, ITC, Dabur), Auto manufacturers (Tata Motors, Mahindra), Retail chains (Reliance Retail, DMart), Real estate developers, Pharma companies

Pain point: A company with 500 vendors needs to manually reconcile what each vendor filed vs. what they claimed as ITC. One mismatch = ITC reversal = cash loss.

Willingness to pay: ₹5L–₹50L/year as enterprise SaaS

Secondary Customers (Government / Institutional)

4. GSTN Itself (Government of India)

The GST Network processes 15 crore+ invoices monthly and loses ₹1.5L Cr+ annually to ITC fraud.

Why they buy: Your knowledge graph does exactly what their internal systems struggle with — real-time circular fraud detection across the entire network.

Engagement model: Tender/RFP, pilot program, or acquisition

5. State GST Departments

Each state has its own GST enforcement wing that investigates fraud.

Who they are: Maharashtra GST, Karnataka Commercial Taxes, UP GST Department

Why they buy: They need investigative tools to audit high-risk taxpayers. A graph that visualizes ITC chains is far more powerful than spreadsheet analysis.

Engagement model: Government procurement / GeM portal listing

6. Banks & NBFCs (For Credit Decisioning)

Banks increasingly use GST data to assess business loan eligibility.

Who they are: SBI, HDFC Bank, Axis Bank, Bajaj Finserv, Capital Float

Why they buy: A business with fraudulent ITC patterns is a credit risk. Your graph gives them a risk score on any GSTIN before they lend money.

Willingness to pay: ₹10L–₹1Cr/year for API access

Tertiary Customers (Indirect / Emerging)

7. E-Commerce Marketplaces

Platforms that onboard lakhs of sellers and are responsible for TCS (Tax Collected at Source).

Who they are: Amazon India, Flipkart, Meesho, Nykaa

Why they buy: They need to verify seller GSTINs are legitimate and not using the platform for fake invoice circulation.

8. Insurance Companies

Trade credit insurance and invoice discounting companies need to verify invoice authenticity.

Who they are: Euler Hermes, ICICI Lombard trade credit, invoice discounting fintechs like KredX, M1xchange

Why they buy: Fake invoices used for financing is a massive fraud vector. Graph verification protects them.

9. Legal & Forensic Accounting Firms

Used during corporate acquisitions, due diligence, or litigation.

Who they are: Forensic teams at EY, Grant Thornton, SBI Cap Securities

Why they buy: When acquiring a company, buyers need to verify the target's ITC history is clean. One bad audit notice can kill a deal.

Customer Priority Matrix

Customer	Market Size	Willingness to Pay	Sales Cycle	Priority
GST Suvidha Providers	Small (30 GSPs)	Very High	Long	 Highest
Large Enterprises	Medium (5000+)	High	Medium	 High
CA Firms	Large (1.4L firms)	Medium	Short	 High
Banks/NBFCs	Small (50 key)	Very High	Long	 Medium
State GST Depts	28 states	Medium	Very Long	 Medium
GSTN	1	Massive	Very Long	 Strategic

7.3 Revenue Projections

Year	Paid Users (B2C)	B2B Contracts	ARR Estimate	Key Milestone
Year 1 (2025–26)	5,000	10 CAs, 1 bank pilot	₹1.2 Cr	Product-market fit, WhatsApp bot live
Year 2 (2026–27)	50,000	100 CAs, 5 banks, 2 NBFCs	₹14 Cr	Series A, 3 language support
Year 3 (2027–28)	3,00,000	500 CAs, 20 banks, enterprise deals	₹85 Cr	GSTN partnership, pan India scale

7.4 Unit Economics

- Customer Acquisition Cost (CAC): ₹200–400 (via WhatsApp organic + accountant referrals)
- Average Revenue Per User (ARPU): ₹2,400/year (Business plan)
- Gross Margin: ~78% (AI API costs are the primary variable cost)
- LTV:CAC Ratio (target Year 2): 12:1
- Payback Period: 2–3 months

8. Target Audience & Go-to-Market

8.1 Primary Personas

Persona 1: Ramesh — Kirana Store Owner, Nashik

Age 38 | Education: 10th pass | Monthly turnover: ₹3–5 lakh | Tech: WhatsApp power user, no laptop Pain: Spends ₹500/month on a local accountant just to file GST. Gets notices occasionally. Doesn't understand the portal. Needs: Send invoice photo on WhatsApp, get confirmation. Nothing more.

Persona 2: Priya — Freelance UI Designer, Bengaluru

Age 27 | Education: B.Des | Annual income: ₹18 lakh | Tech: iPhone, Figma, Instagram Pain: Has to file both GST (for design services) and ITR. Pays ₹8,000/year to a CA. Invested only ₹60K in ELSS out of ₹1.5L limit. Needs: One app that handles both GST and tells her exactly where to invest before March 31.

Persona 3: Vikram — Chartered Accountant, Ahmedabad

Age 45 | Manages 200+ clients | Staff: 4 assistants | Pain: Data collection from clients is chaotic — WhatsApp photos, emails, PDFs Needs: A tool that lets clients upload invoices themselves, auto-parses them, and presents clean data for review. Willing to pay ₹3,000/month for this.

8.2 Go-to-Market Strategy

Phase 1 (Months 1–6): Grassroots Launch

- Launch WhatsApp bot in Maharashtra (Hindi) targeting Kirana stores via local business WhatsApp groups
- Partner with 10–15 Chartered Accountants who give access to their existing client base in exchange for free Pro accounts
- Launch on Product Hunt and LinkedIn to attract early tech adopters and freelancers
- Target: 500 active users, 3 CA partnerships

Phase 2 (Months 7–12): Channel Expansion

- Integrate with 2 regional banks (cooperative banks in Maharashtra, Gujarat) for MSME loan cross-sell
- Partner with e-commerce platforms (Meesho, Flipkart Seller Hub) whose sellers need GST compliance tools

- Add Tamil and Telugu language support; partner with trade associations in Chennai and Hyderabad
- Target: 5,000 paid users, 50 CA accounts, 1 bank partnership

Phase 3 (Year 2): B2B Scale

- Launch Enterprise Payroll module with 3 mid-size companies (500–5,000 employees) as pilot
- Approach GSTN / CBIC for pilot partnership on fraud analytics data sharing
- Series A fundraise targeting ₹15–20 Cr at ₹80–100 Cr valuation
- Expand to 5 languages and 10 states

9. Business Strategies

9.1 Growth Strategy: The Accountant Flywheel

The core distribution insight is that Chartered Accountants are the trust layer for 80% of Indian MSMEs. If a CA recommends TaxIQ to their 200 clients, that's 200 instant users with high trust conversion. We build a best-in-class CA Pro tier (₹3,000/month) that CAs love — and CAs become our most powerful sales channel at near-zero CAC.

9.2 Retention Strategy: The Tax Calendar Loop

Unlike most apps that are used once a year at tax time, TaxIQ is designed for monthly engagement through three hooks:

- Monthly GST filing reminder + one-click re-file if nothing changed
- Monthly '80C Progress Report' showing investment vs. target with months remaining
- Real-time fraud alerts: 'One of your suppliers has a suspicious ITC pattern — review before next filing'

9.3 Data Strategy: The Knowledge Graph Moat

Every invoice processed builds a denser, more accurate fraud detection network. By Month 12 with 5,000 users, we will have processed an estimated 200,000+ invoices and mapped 50,000+ GSTIN relationships — creating a proprietary fraud signal network that no new entrant can replicate quickly. This graph becomes a licensable asset for banks, NBFCs, and the government.

9.4 Regulatory Strategy

- Register as a GST Suvidha Provider (GSP) to get official API access to GSTN
- Engage with the CBIC Sandbox program for testing and validation
- Build a compliance-first culture: all user data stays in India (AWS Mumbai), SOC 2 certified by Year 2
- Participate in DPIIT Startup India program for tax benefits and government pilot access

10. Financial Plan & Funding

10.1 Cost Structure (Year 1)

Cost Category	Monthly	Annual	Notes
Engineering (2 devs + 1 ML)	₹3,00,000	₹36,00,000	Senior full-stack + ML engineer + junior
LLM API (Claude)	₹15,000	₹1,80,000	Scales with invoice volume
Google Cloud Vision	₹5,000	₹60,000	OCR costs
Neo4j AuraDB	₹8,000	₹96,000	Graph database hosting
Supabase + hosting	₹10,000	₹1,20,000	Database + deployment
Twilio (WhatsApp)	₹12,000	₹1,44,000	Message costs at 10K users
Marketing + partnerships	₹50,000	₹6,00,000	Accountant outreach, digital ads
Legal + compliance	₹20,000	₹2,40,000	GSP registration, T&C, privacy
Miscellaneous	₹10,000	₹1,20,000	Tools, subscriptions
TOTAL	₹4,30,000	₹51,60,000	

10.2 Funding Ask

Seeking: ₹75 Lakhs Pre-Seed Round Use of Funds: 48% Engineering | 28% Marketing & Partnerships | 14% Legal & Compliance | 10% Operations Valuation: ₹3.75 Cr pre-money (20% equity) Runway: 18 months to Series A milestones

11. Team Requirements

Role	Skills Required	Hire Priority
Full Stack Engineer (Lead)	Python, FastAPI, React/Streamlit, PostgreSQL, Docker	Day 1
ML / AI Engineer	LangChain, Claude API, OCR pipelines, Neo4j Cypher, Pandas	Day 1
Product Designer	Mobile-first UX, low-literacy design, WhatsApp chatbot flows	Month 2
Tax Domain Expert (Advisor)	GST filing, ITC rules, Income Tax sections, CA background	Month 1 (part-time)
Business Development	CA network, MSME associations, bank partnerships	Month 3
Junior Developer	Frontend, testing, deployment automation	Month 4

12. Risk Analysis & Mitigation

Risk	Probability	Impact	Mitigation
GSTN API access denial / GSP registration delay	Medium	High	Launch with mock GSTN responses; fast-follow with GSP application; partner with existing GSPs as interim solution
LLM parsing errors on complex invoices	Medium	Medium	Human-in-the-loop review for first 3 months; confidence scoring rejects low-confidence parses for manual review
ClearTax launches similar AI feature	High	Medium	Speed to market + WhatsApp moat + knowledge graph depth are hard to replicate quickly; focus on segments ClearTax ignores
Data privacy concerns from users	Medium	High	India data residency, ISO 27001 certification, transparent privacy policy, no selling of financial data, regular audits
Low digital literacy of target MSME segment	High	Medium	WhatsApp-first design requires zero learning curve; vernacular voice explanations; in-person onboarding via CA partners
Tax law changes (e.g., 80C restructuring)	Low	High	Tax rules engine is modular and independently updatable without engineering deploys; monitor CBDT circulars

13. Vibe Coding Prompt

Use the following prompt to bootstrap TaxIQ with any AI coding assistant (Claude, Cursor, v0, etc.):

```
# TaxIQ — Vibe Coding Master Prompt

Build TaxIQ — India's unified AI tax intelligence platform.
Tech stack: Python + FastAPI backend, Streamlit frontend,
Anthropic Claude API for AI, Neo4j for knowledge graph,
PostgreSQL (Supabase) for data, LangChain for agents.

## AGENT 1: GST Invoice Filing Agent
- Accept invoice image (JPG/PNG/PDF) via file upload or URL
- Use pytesseract for OCR to extract raw text from the image
- Pass image + raw text to Claude API with prompt:
  "Extract: vendor_gstin, buyer_gstin, invoice_number,
  invoice_date, hsn_code, taxable_value, cgst, sgst,
  igst, total. Return ONLY valid JSON."
- Validate: GSTIN format (15 chars), tax math (cgst+sgst+igst)
- Store in PostgreSQL invoices table
- Build GSTR-1 JSON payload from multiple invoices
- Show parsed invoice in Streamlit with edit capability

## AGENT 2: Knowledge Graph Fraud Detection Agent
- Connect to Neo4j (local bolt://localhost:7687)
- For each invoice: create GSTIN nodes, Invoice node,
  ISSUED and CLAIMED_ITC_FROM relationships
- Run Cypher cycle detection query:
  MATCH path=(a:GSTIN)-[:CLAIMED_ITC_FROM*3..6]->(a)
  RETURN path LIMIT 10
- Calculate risk_score: 0=safe, 1=high risk
- Display interactive pyvis graph in Streamlit
- Show Compliance Confidence Score (0-100) prominently
- Load mock_gstin_data.json with 3 fraud scenarios:
  (1) circular chain, (2) star pattern, (3) clean network

## AGENT 3: Personal Tax Optimizer Agent
```

- Accept bank statement CSV (any Indian bank format)
- Use Pandas to normalize: date, description, debit, credit
- Classify transactions using keyword matching + Claude:
LIC/insurance -> 80D | ELSS/mutual fund -> 80C
PPF/NSC -> 80C | home loan EMI -> split 80C + 24B
NPS -> 80CCD(1B) | school fees -> 80C
- Tax limits: 80C=150000, 80D=25000, 80CCD1B=50000, 24B=200000
- Show gap table: [section | limit | invested | gap | tax_saved]
- Compare Old vs New regime with actual calculated numbers
- Call Claude to generate Hindi + English recommendation text
- Generate downloadable PDF report using reportlab

STREAMLIT DASHBOARD STRUCTURE

Page 1: 'GST Filing' – upload invoice, show parsed JSON, editable fields, Add to GSTR-1 button, filing summary
 Page 2: 'Fraud Dashboard' – pyvis graph, risk scores, Compliance Confidence Score, flagged GSTINs table
 Page 3: 'Tax Saver' – CSV upload, gap analysis table, regime comparator, AI recommendations, PDF download

SAMPLE DATA TO CREATE

1. data/sample_invoices/ – 5 invoice JPGs (create mock images with realistic Indian GST invoice data using Pillow)
2. data/sample_bank_statements/hdfc_statement.csv – 6 months of transactions with LIC, SIP, medical bills, school fees
3. data/mock_gstn_data/fraud_scenarios.json – 3 scenarios: circular chain (5 nodes), star fraud, clean network

ENVIRONMENT SETUP

Create .env with: ANTHROPIC_API_KEY, NEO4J_URI, NEO4J_USERNAME, NEO4J_PASSWORD, DATABASE_URL
 Create requirements.txt, docker-compose.yml with postgres + neo4j
 Create README.md with setup steps and demo walkthrough

Build everything as a single working repo. Prioritize:
 (1) working demo over perfect code
 (2) real data parsing over hardcoded examples
 (3) visual impact on the Streamlit dashboard
 (4) bilingual (Hindi + English) output from Claude

14. Product Roadmap

Quarter	Milestone	Feature Focus	Success Metric
Q1 2025	MVP Launch	GST Invoice Agent + Streamlit dashboard (Hindi + English)	500 invoices processed
Q2 2025	WhatsApp Beta	WhatsApp Bot via Twilio + Fraud Graph (5 patterns)	200 active WhatsApp users
Q3 2025	Tax Optimizer Live	Bank CSV parser + 80C/D gap analysis + regime comparator	1,000 paid users
Q4 2025	CA Pro Tier	CA dashboard: multi-client view, bulk invoice processing	20 CA accounts
Q1 2026	Bank Integration	API licensing for MSME credit scoring + Tamil/Telugu	2 bank partnerships
Q2 2026	Series A	GSTN partnership pilot + 3 enterprise payroll accounts	₹14 Cr ARR trajectory
Q4 2026	Platform Scale	10 languages, 5 states, white-label offering	3,00,000 users

TaxIQ

India's Unified Tax Intelligence Agent

From Invoice Photo to ITR — One AI Platform