

**SHRI GOVINDRAM SEKSARIA INSTITUTE OF  
TECHNOLOGY & SCIENCE INDORE MADHYA  
PRADESH**

**A**

**MAJOR PROJECT REPORT : PHASE-1**



**BACHELOR'S OF TECHNOLOGY  
IN  
COMPUTER SCIENCE ENGINEERING**

**Project Title -- TaxES**

**Your Names --** Payal Badal (0801CS221107)

Sameeksha Suryawanshi (0801CS221122)

Shivank Rai (0801CS221130)

Stuti Badola (0801CS221139)

Yogesh Sahu (0801CS233D12)

**Supervisor's Name --** Associate Professor Surendra Gupta

**Date of Submission --** 25/03/2025

DEPARTMENT OF COMPUTER ENGINEERING SHRI G.S.  
INSTITUTE OF TECHNOLOGY & SCIENCE INDORE

**DURATION :** 6th JAN 2025 – 24th MARCH 2025

## ABSTRACT

### Brief Summary of the Project

#### Key Objectives

The TaxES project aims to simplify and automate the income tax filing process for Indian individuals by leveraging expert systems and natural language processing (NLP). The objective is to reduce the complexity of tax calculations, minimize manual data entry, and ensure accurate filing by extracting key financial details from documents like salary slips and bank statements. By integrating NLP, OCR, and AI-based classification, the system can identify relevant information, compute taxable income, and suggest the correct ITR form for the user.

#### Main Outcomes-

**Automatic Data Extraction** – The system reads and collects important financial details from salary slips, bank statements, and other documents without needing manual entry.

**Pre-filled Tax Forms** – Users get tax forms that are already filled with their details, making filing much easier.

**Smart Tax Calculation** – The system calculates taxable income and deductions automatically, ensuring accuracy.

**Tax Form Recommendation** – Based on the user's financial data, the system suggests the correct ITR form to use.

**Simple and User-friendly Process** – People without tax knowledge can easily file their returns with step-by-step guidance.

## Background and Context

### Motivation of the Project :

The **TaxES** project aims to simplify income tax filing for Indian individuals by automating data extraction and tax calculations. Many taxpayers struggle with complex tax laws, selecting the correct ITR forms, and manual data entry, leading to errors and delays. To address this, **TaxES** leverages AI, expert systems,

NLP, and OCR to extract financial details from salary slips and bank statements, reducing manual effort and ensuring accuracy.

### **Objective :**

The main objective of this project is to create an intelligent tax filing platform that guides users step-by-step, minimizing errors and maximizing efficiency. The system will automatically identify relevant tax deductions, compute taxable income, and recommend the correct ITR form based on individual financial data. Another key objective is to enhance compliance by detecting inconsistencies in tax filings and suggesting corrective actions before submission.

### **Scope :**

The scope of the project includes designing an AI-powered platform that integrates expert systems, NLP models, and OCR technology to automate data extraction, tax calculations, and form filling. The platform will cater to salaried individuals, freelancers, and small business owners who need a hassle-free way to file their taxes. It will provide a secure and user-friendly interface where users can upload financial documents, get real-time tax calculations, and receive personalized filing recommendations. The ultimate goal is to reduce the complexity of tax filing, save time for taxpayers, and improve the accuracy of tax submissions in India.

## **Literature Review**

### **Related Work:**

- A study explored the role of AI chatbots in income tax prediction in India, highlighting their potential to assist individuals and businesses in navigating complex tax systems. **Journal Press India**
- Infosys developed an automated tax computation solution that leverages machine learning to extract data from diverse sources, aiming to improve accuracy and efficiency in tax computations. **Navigate your next**
- Research has demonstrated how AI technologies, including large language models (LLMs), NLP, and machine learning, can enhance

tax audit reporting by improving efficiency and accuracy in tax compliance processes. **ResearchGate**

### **Existing Income Tax e-Filing Solutions:**

#### **Income Tax e-Filing Portal**

- Official government portal for tax filing in India.
- Provides access to tax laws, various ITR forms, and an Income Tax Calculator.

#### **Vakil Search**

- Offers expert legal and compliance solutions for tax filing and other financial matters.

#### **MyTaxIndia**

- Focuses on individual tax calculations, helping taxpayers estimate their tax liabilities.

#### **ClearTax**

- Provides tax calculation, return filing, and additional financial services.
- Uses automation to simplify tax filing for individuals and businesses.

### **Gaps in Existing Knowledge:**

Despite these advancements, challenges remain in fully automating the tax filing process for individuals, particularly in accurately extracting and interpreting unstructured financial data from various document formats. Additionally, integrating these technologies into a seamless, user-friendly platform tailored to the specific requirements of Indian taxpayers is an area that requires further research and development.

## **Problem Statement :**

*“How might we simplify and automate the process of filing income tax returns for Indian individuals by using expert systems and NLP to ensure accuracy and reduce complexity?”*

## **Methodology for TaxES Project**

### **Detailed Plan of Action**

- **Data Collection & Preprocessing:**

Extract financial details from salary slips, Form 16, and bank statements using **OCR**.

- **Natural Language Processing (NLP):**

Identify relevant tax-related information and classify income sources.

- **Expert System Development:**

Automate tax calculations, deductions, and suggest appropriate **ITR forms**.

- **Platform Development:**

Build a user-friendly web and mobile interface for seamless interaction.

- **Testing & Validation:**

Conduct accuracy tests on extracted and processed tax data.

### **Experimental or Theoretical Framework**

- **OCR for Text Extraction:**

Extracts financial data from scanned documents.

- **NLP-based Classification:**

Identifies and categorizes financial terms.

- **Expert System Logic:**

Applies Indian tax rules to compute tax liability.

- **Machine Learning (ML):**

Enhances accuracy in tax classification and deductions.

## **Tools and Technologies for TaxES AI**

### **Artificial Intelligence & Machine Learning**

- **OCR (Optical Character Recognition):**  
Tesseract OCR, Google Vision API.
- **Natural Language Processing (NLP):**  
spaCy, BERT, OpenAI GPT, Hugging Face Transformers.
- **Machine Learning Models:**  
Scikit-learn, TensorFlow, PyTorch.
- **Expert System Development:**  
Rule-based AI using Drools or Prolog.

## **Application Development**

- **Backend:** Django, FastAPI for high-performance API handling.
- **Frontend:** React Native for mobile, Next.js for web UI.
- **Database:** PostgreSQL or MySQL for structured tax data storage.
- **Authentication:** OAuth 2.0, JWT for secure access.

## **Cloud & Deployment**

- **Hosting:** AWS, Google Cloud, Play Store or Azure for scalable deployment.
- **Containerization:** Docker, Kubernetes for microservices architecture.
- **CI/CD Pipeline:** GitHub Actions, Jenkins for automated deployment.

## **APIs & Integrations**

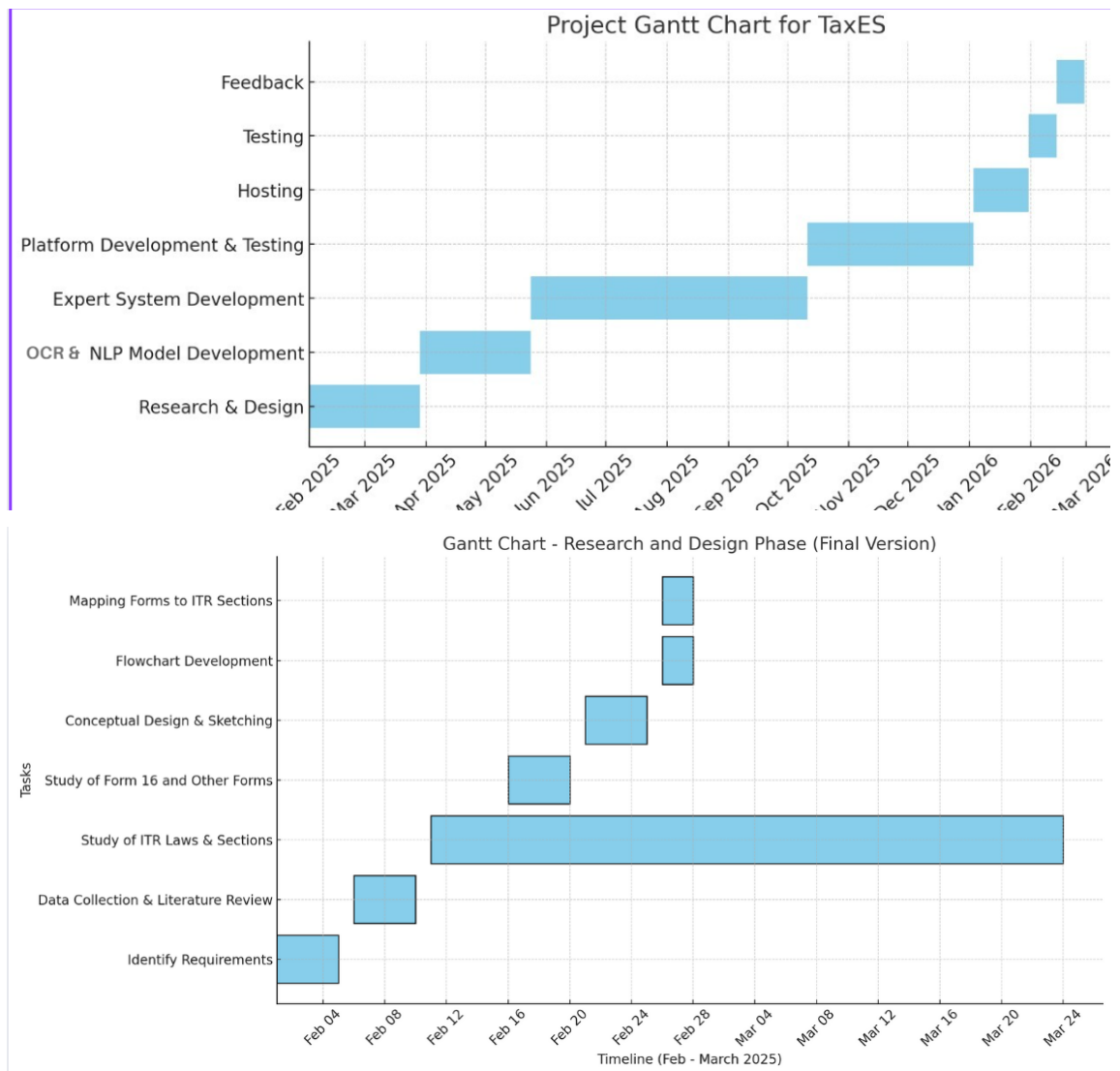
- **Income Tax Department API:** For tax validation and compliance.
- **Banking & Financial APIs:** For direct financial data retrieval.

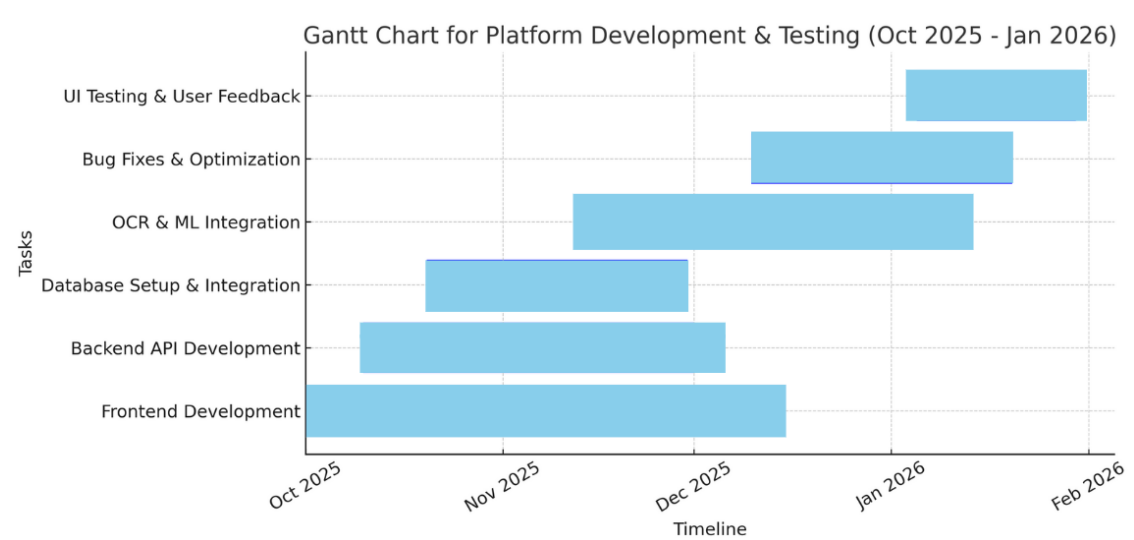
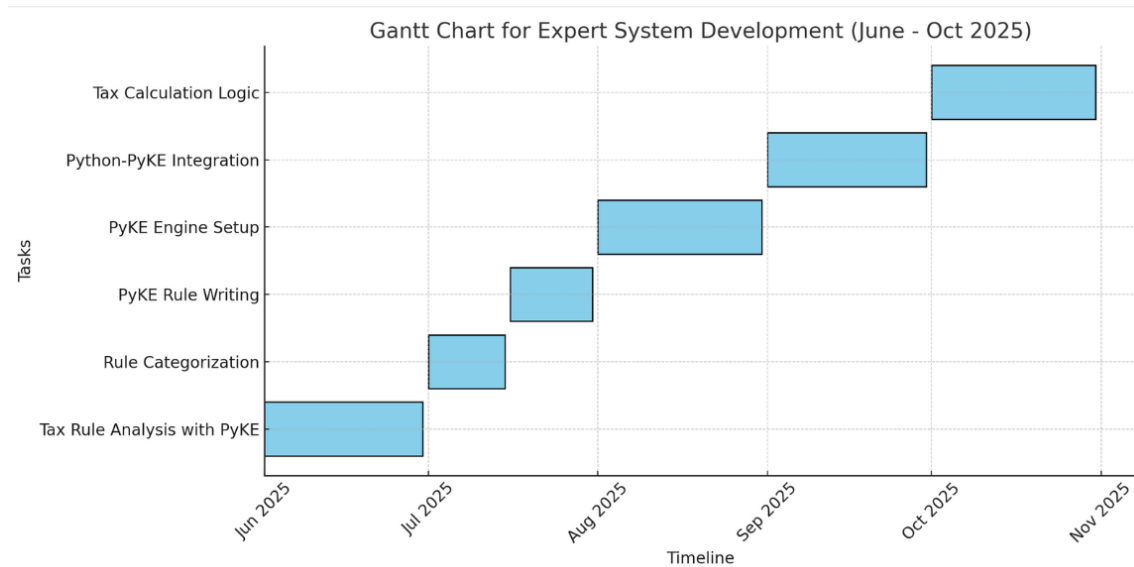
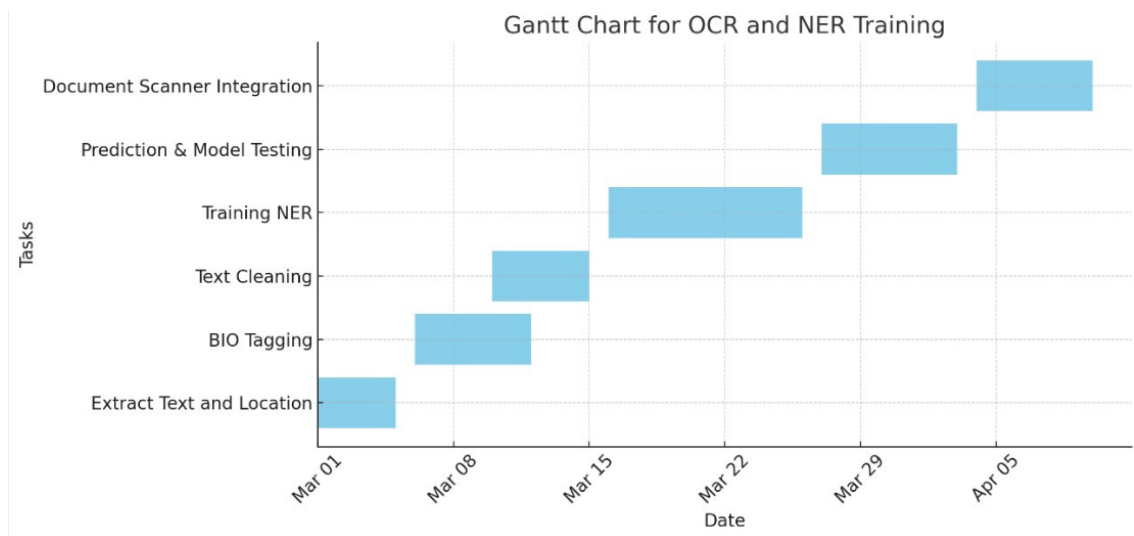
- **AI-based Document Processing:** Google Document AI for form analysis.

## Testing & Security

- **Unit & Integration Testing:** PyTest, Jest for reliable application testing.
- **Security Tools:** OWASP ZAP, SonarQube for vulnerability assessment.

## Project Timeline







Module	Start Date	End	Member Assigned
Requirement Gathering & Analysis - Research & Design	Feb-25	Mar-25	Shivank Rai, Stuti Badola, Sameeksha Suryawanshi, Payal Badal
NLP Model Development	Apr-25	Jun-25	Yogesh Sahu, Shivank Rai, Stuti Badola
Expert System Development	Jul-25	Sep-25	Yogesh Sahu, Stuti Badola, Payal Badal
Platform Development & Testing	Oct-25	Dec-25	Stuti Badola, Sameeksha Suryawanshi, Payal Badal
Hosting & Deployment	Jan-26	Jan-26	Shivank Rai, Yogesh Sahu
Testing	Feb-26	Feb-26	Yogesh Sahu, Shivank Rai, Stuti Badola, Sameeksha Suryawanshi, Payal Badal
Feedback	Mar-26	Mar-26	Yogesh Sahu, Shivank Rai, Stuti Badola, Sameeksha Suryawanshi, Payal Badal

## Resources Required

### Software & Datasets

- **Programming Languages:** Python, JavaScript for AI model development and application interface.
- **Machine Learning Frameworks:** TensorFlow, PyTorch for training and deploying AI models.
- **NLP Libraries:** spaCy, BERT, OpenAI GPT for processing financial documents.
- **OCR Tools:** Tesseract OCR, Google Vision API for extracting text from documents.
- **Databases:** PostgreSQL, Firebase for storing tax-related data.
- **Cloud Services:** AWS, Google Cloud, or Azure for hosting AI models and application backend.
- **APIs:** Income Tax Department API, Banking APIs for real-time data access.
- **Testing Tools:** Selenium, PyTest for ensuring software quality.

### Hardware Requirements

- **High-performance servers** for AI model training and deployment.
- **GPUs (Graphics Processing Units)** for deep learning model acceleration.
- **Secure cloud storage** for handling user tax-related documents safely.

## Reference

**Income Tax e-Filing Portal:** The official government platform that provides access to tax laws, various forms, and features an Income Tax Calculator.

<https://www.incometax.gov.in/iec/foportal/>

**Vakil Search:** Offers expert legal and compliance solutions, assisting individuals and businesses with their tax filing needs.

<https://vakilsearch.com/>

**MyTaxIndia:** Provides tax calculation services tailored for individuals, simplifying the process of determining tax liabilities.

<https://mytaxindia.com/>

**ClearTax:** Offers comprehensive services including tax calculation, return filing, and additional financial services to streamline the tax filing process.

<https://cleartax.in/>

[https://www.researchgate.net/publication/389144454\\_AI-Driven\\_Tax\\_Technology\\_in\\_the\\_United\\_States\\_A\\_Business\\_Analytics\\_Framework\\_for\\_Compliance\\_and\\_Efficiency](https://www.researchgate.net/publication/389144454_AI-Driven_Tax_Technology_in_the_United_States_A_Business_Analytics_Framework_for_Compliance_and_Efficiency)

<https://tax.thomsonreuters.com/blog/the-impact-of-artificial-intelligence-on-the-tax-and-accounting-profession/>