

# Bhanuprasadh Santra

📍 Bengaluru, India    ✉ bhanuprasadh004@gmail.com    ☎ 9701152083    🔗 LinkedIn    🐙 GitHub

## EXPERIENCE

### Software Engineering Intern, Xylon Research

Jun 2025 | Bangalore, India

- Automated tax compliance analysis using AI models to detect regulatory inconsistencies in financial documents, cutting manual review time by 70%.
- Developed a multi-LLM architecture (GPT-4o + Groq's kimi-k2-instruct) to cross-verify invoice data extraction results, boosting accuracy by 40%.
- Built and deployed a FastAPI-based backend for all compliance tools, ensuring high concurrency, secure authentication, and smooth scalability.
- Designed an automated validation engine that translates complex UAE FTA VAT laws into structured, programmable rules.
- Implemented a resilient data pipeline for CSV invoice ingestion, integrating schema validation, detailed error logs, and automated cleanup scripts.

## PROJECTS

### Local RAG Document Assistant

Nov 2025

- Architected a local Retrieval-Augmented Generation (RAG) pipeline using LangChain and ChromaDB to enable accurate, context-aware Q&A on private documents.
- Implemented semantic search capabilities using Sentence Transformers and integrated Llama 3 (via Ollama) for secure, offline LLM inference without data leakage.
- Optimized the retrieval system by fine-tuning text splitting strategies and embedding models to significantly improve the relevance and accuracy of generated responses.

### Gemini-Powered RAG Knowledge Bot

Nov 2025

- Engineered a full-stack RAG solution for contextual data querying, achieving high accuracy by integrating the Google Gemini LLM with a proprietary knowledge base.
- Optimized data retrieval using ChromaDB for efficient vector storage and similarity search. Leveraged LlamaIndex to manage indexing and minimize query latency.
- Deployed a production-ready application with an interactive Streamlit frontend, ensuring a seamless user experience with robust conversational context.

### Network Security: AI-Powered Phishing Detection System

Nov 2025

- Engineered an end-to-end Machine Learning pipeline for phishing detection using Scikit-learn, implementing custom data transformation and validation components to ensure high model accuracy.
- Developed a high-performance REST API using FastAPI and MongoDB to serve real-time predictions, optimizing data processing workflows for large-scale network security datasets.
- Designed a modular and scalable code architecture, utilizing object-oriented programming principles and robust exception handling to enhance system maintainability and reliability.

### Advanced Web Scraper API

Oct 2025

- Developed a production-ready web scraping API using Python and FastAPI, supporting static and dynamic sites through BeautifulSoup and Playwright.
- Designed the system to handle bulk scraping, data extraction, and screenshot capture, with built-in error handling, logging, and concurrency control for reliable performance.
- Packaged the application into a Dockerized microservice with RESTful endpoints, health checks, and Swagger documentation for easy deployment and integration.

## SKILLS

**Languages:** Python, Go, SQL, Bash/Shell Scripting

**Backend & APIs:** FastAPI, Flask, RESTful APIs, Microservices Architecture

**AI/ML Frameworks:** PyTorch, TensorFlow/Keras, Scikit-learn, XGBoost, Hugging Face

**LLMs & Prompt Engineering:** LangChain, ChromaDB, Ollama, Vector Search, LLMs (Llama 3), RAG, Prompt Engineering

**Tools:** Docker, Kubernetes, AWS (S3, EC2, Lambda), CI/CD (GitHub Actions), MLFlow, Git/GitHub, Dagshub

**Databases:** MySQL, MongoDB, PostgreSQL

## EDUCATION

### Kakinada Institute of Engineering and Technology,

Nov 2021 – Apr 2025 | Kakinada, India

B.Tech in Computer Science (AIML)

**Relevant Coursework:** Data Structures & Algorithms, Database Systems, Operating Systems, Computer Networks, Distributed Systems, Machine Learning, Deep Learning