

Yohan Vergis Vinu

yohanvv@seas.upenn.edu | +91 75500 26048 | [LinkedIn](#) | [GitHub](#)

EDUCATION

University of Pennsylvania | School of Engineering and Applied Science
Master of Science in Engineering in Computer and Information Science
Incoming Student, Fall 2025

Philadelphia, PA
Expected May 2027

SRM Institute of Science and Technology | School of Computing
Bachelor of Technology in Computer Science and Engineering | **GPA: 9.80/10**

Chennai, India
June 2025

Coursework: Intro to ML, Digital Image Processing, Data Mining and Analytics, AI, Database Management Systems

Extracurricular: Class Representative (8 consecutive semesters); IoT Lead & Founding Member, Networking Nexus SRM

EXPERIENCE

Moative

Houston, TX | Remote

AI/ML Engineer Intern

Aug. 2024 – Present

- Spearheaded end-to-end development of [Vina](#), an AI-powered CHRO platform that equips HR with ideal traits and questions, automates candidate analysis and comparison using multimodal agentic architecture on a Django/PostgreSQL AWS stack
- Built FastAPI endpoints with RAG-powered AI agents that automate company research through web crawling, cutting manual lookup times by 75%
- Developed IDP workflows with fine-tuned VLMs to enhance financial data extraction across loan and insurance documentation
- Leading development of Leslie, an agentic AI assistant that streamlines government permitting and licensing in Rowlett, TX. Built with Django, LangGraph, and VLMs, the system reduces permit issuance wait times from weeks to minutes

Samsung R&D Institute India

Chennai, India

ML Research Intern

Aug. 2023 – March 2024

- Led a team of four to develop a fine-tuned MobileNetV2-SSD model for detecting specific features of cats and dogs across 9 classes (e.g., eyes, ears, mouth) and optimized the model for on-device purposes
- Compiled and annotated a custom dataset of over 10,000 images, covering over 50 dog breeds and 15 cat breeds
- Achieved a 5-11% improvement in accuracy over existing methods, earning **Samsung's Certificate of Excellence**

PROJECTS

Ensemble Models for Personal Expenditure Forecasting

April 2025

- Analyzed DL models (LSTM, N-BEATS) for forecasting personal expenditures and developed a custom ensemble model, achieving a 4–8% improvement in MASE over individual baselines
- Designed a graph-based synthetic financial data generator powered by LLMs, capable of producing customizable datasets spanning 24 invoice attributes across 37 categories in minutes
- Authored (in preparation) a research paper titled “*Synthetic Data Generation and Neural Ensemble Architectures for Personal Expenditure Forecasting*”, exploring model interpretability and predictive accuracy

InvoiceGPT: Agentic Invoice Processing

Dec. 2024

- Architected an AI-driven invoice processing platform using GPT-4 Vision for OCR, and a custom algorithm for accurate line-item segmentation, deployed on AWS infrastructure (S3, RDS) with Google OAuth for robust data isolation
- Built a LangGraph-powered conversational interface enabling natural language queries over user-specific financial data
- Developed an interactive analytics dashboard to visualize spending patterns and forecast expenditures

GRE Verbal Question Generator

Aug. 2024

- Built an AI system using LLaMA 3.1, RAG, and ChromaDB to generate GRE-style verbal questions from concept-level prompts
- Engineered a retrieval pipeline with OpenAI embeddings and topic-aware metadata for accurate, context-aligned MCQ generation

ACHIEVEMENTS

Internal Round Finalist (Top 25/400+ teams), Smart India Hackathon 2023, SRM, India

Sept. 2023

- Developed a prototype of an all-in-one multilingual application for farmers, including crop price prediction & recommendation, weather alerts, cold-storage information, market access transportation, tool marketplace, and chat-bot assistance

7th Position (Out of 130+ teams), Kodekshetra Hackathon, Think-Digital, SRM, India

April 2023

- Engineered a lightweight facial obstruction detection framework for enhanced ATM security using Haar classifiers and computer vision. The purpose is to make the ATM inoperable until the user is completely identifiable

SKILLS

Programming & Scripting: Python, SQL, C/C++, Java, JavaScript

AI/ML: PyTorch, TensorFlow, HuggingFace, LangChain/LangGraph, RAG, OpenCV, VLMs, NumPy, Pandas, Tableau, scikit-learn

Web Development & APIs: Django, Flask, FastAPI, Streamlit, Tailwind CSS, RESTful APIs, Postman

Cloud, Databases & DevOps: AWS, GCP, Azure, PostgreSQL, MySQL, Pinecone, ChromaDB, Docker, Git/GitHub

CERTIFICATIONS

Deep Learning Specialization, DeepLearning.AI and Stanford University

April 2024 - Aug. 2024

Machine Learning Specialization, DeepLearning.AI and Stanford University

Sept. 2023 - Feb. 2024