

# Pragathi Kalidasan Vetrivel Murugan

Chennai, IN | P: +91 95008 57575 | [Email](#) | [LinkedIn](#) | [Github](#) | [Portfolio](#)

## PROFILE SUMMARY

Associate Data Scientist with 2 years of experience focused on the development and production deployment of AI solutions. Experienced in the full Large Language Model (LLM) lifecycle, including hands-on practice pre-training models from scratch. Demonstrated experience building modular data pipelines using Kedro to transition experimental research into maintainable, production-ready systems.

## CORE SKILLS

- **Programming:** Python, C++, SQL
- **Frameworks & Tools:** Kedro, Trino, Streamlit, LangChain, Ollama, PyTorch, Scikit-learn, FastAPI, GCP
- **Skills:** LLMs, SLMs, Data Pipelines, CI/CD, Dashboard Development, Machine Learning, Quantitative Finance, Reinforcement Learning, Vibe Coding

## WORK EXPERIENCE

### SOCIÉTÉ GÉNÉRALE GLOBAL SOLUTION CENTRE

Bengaluru, IN

#### Associate Data Scientist

Aug 2024 – Present

- Built production-grade data pipelines and implemented CI/CD workflows for deployment
- Developed full stack AI applications including React and Streamlit for interactive data solutions
- Contributed to Generative AI initiatives, leveraging LLMs for automation and intelligent reporting

#### Key Projects:

- **Controller's Assistant:** An organization wide initiative to transform processes and adopt AI usage among LoD2 controllers
  - **Review Assistant:** Built two data pipelines using Kedro to process raw data and convert it into comprehensive data products, which enabled streamlined review of activities, risk identification and control assessments. This solution empowered controllers to efficiently pinpoint gaps in risk and control frameworks, conduct targeted reviews, and enhance overall process effectiveness
  - **Report Automation:** Deployed an end-to-end AI solution leveraging advanced prompt engineering workflows to automate and refine restitution report findings, reducing manual review cycles by 85-90% between controllers and supervisors
- **Predictive Data Completeness:** An initiative aimed at improving the data quality of the Group Referential data and improve the operational processes revolving around it
  - **Data remediation:** Built a two-phase approach combining rule-based automation and XGBoost ML model to remediate missing data in the group's three most critical attributes, improving data quality. Deployed the final models as APIs for direct consumption by the analysts
  - **Data analysis:** Conducted a comprehensive analysis of the group's most critical attributes, uncovered anomalies not addressed by existing controls, and proposed new BCBS239-compliant controls, significantly enhancing data governance and risk mitigation
- **Cross-functional AI Mentor:** Collaborated with Data Scientists from other team to develop an AI-based onboarding assistant for new joiners

### ROOTQUOTIENT PRIVATE LIMITED.

Chennai, IN

#### Software Development Engineer Intern

Jun 2022 – Dec 2022

- Deployed and migrated core API services to modernize a legacy Ruby on Rails monolith, improving the interface experience for **15,000+ active users**
- Architected the backend infrastructure for a digital-first loan platform from MVP to public launch, integrating **Ping (Auth)** and **Leegality (e-Sign)** to achieve a 100% paperless loan lifecycle for the initial **100+ active users**.
- Facilitated production deployments and timely fix of bugs

## PROJECTS

---

### SmolLM2-135M

Jan 2026

- Reverse-engineered a 135M-parameter Llama-style transformer from scratch using PyTorch, implementing RoPE, Grouped Query Attention, SwiGLU MLP, and RMSNorm
- Trained the model on custom text data with bfloat16 mixed precision, Flash Attention, gradient accumulation, and cosine learning rate scheduling
- Developed an interactive Gradio web app for text generation with configurable sampling parameters (temperature, top-k, top-p)

### AI Analyst

Sep 2025

- Engineered a full-stack AI Analyst platform that automates the generation of 12-section investment reports from PDF pitch decks, achieving complex analysis via the Google Gemini/Vertex AI suite
- Designed and implemented a modular Retrieval-Augmented Generation (RAG) agent system, utilizing 12 specialized LLM tools for structured analysis in areas such as market opportunity, risk mitigation, and competitive benchmarking
- Developed multimodal data processing capabilities using the Vision API and utilized Google Cloud Storage and Firestore for a scalable, cloud-native architecture capable of producing professional PDF reports and dynamic visualizations

### Option Pricing

Mar 2023

- Implemented and trained a Double Deep Q Network (DQN) agent to learn and execute optimal dynamic hedging strategies, significantly reducing portfolio risk variance for financial options
- Developed a custom financial simulation environment, utilizing the QuantLib library to accurately model stochastic asset dynamics for training the Reinforcement Learning agent
- Integrated the Transformer architecture into the pricing model, enabling the agent to analyze historical price paths for more accurate valuation of complex, path-dependent options

## EXTRA CURRICULAR AND CERTIFICATIONS

---

- Runner up in a hackathon conducted by **Société Générale - August 2025**
- Winner in a hackathon conducted by **Société Générale - August 2024**
- Winner in a whitepaper competition conducted by **Société Générale - August 2024**
- Runner up in a hackathon conducted by **Caterpillar - July 2023**
- **JPMorgan Chase & Co. Quantitative Research Virtual Experience Program on Forage - January 2024**
  - Completed a simulation focused on quantitative research methods
  - Analyzed a book of loans to estimate a customer's probability of default
  - Used dynamic programming to convert FICO scores into categorical data to predict defaults