JOSHUA PETER POLAPRAYIL

Details

Eravimangalam P O
Kottayam, Kerala, 686613
India
+91 8281158864
josh19peter96@gmail.com

Links

GitHub

LinkedIn

Languages

English

Malayalam

Hindi

Tamil

Profile

Junior software engineer with a Master's in Big Data Analytics & AI and hands-on experience in deep learning (CNN, RNN, BioBERT), data pipelines (Airflow, dbt, BigQuery), and fullstack systems (NestJS, TypeScript, React, PostgreSQL). Skilled in deploying applications on AWS/GCP with Docker and CI/CD, and experienced in enterprise development with C#, .NET, and SQL Server. Strong foundation in Python, SQL, and practical problem-solving across AI, data, and product engineering.

Skills

Languages:

Python, TypeScript, C#, SQL, JavaScript, Shell, PowerShell

Artificial Intelligence & Machine Learning:

Generative AI, LLM Integration, Prompt Engineering, Multi-Agent Systems, RAG (Retrieval-Augmented Generation) Deep Learning (CNN, RNN, Bi-LSTM, Transformers), NLP, Computer Vision (Object Detection, Tracking), Machine Learning LangChain, LangGraph, PyTorch, TensorFlow, Google Gemma API, ChromaDB, YOLO, OpenCV Model Optimization, Hyperparameter Tuning, Text Generation, Multi-Modal AI

Data Engineering & Analytics:

Apache Spark, Kafka, Airflow, dbt, ETL Pipelines, Data Manipulation (NumPy, Pandas) Google Cloud Storage (GCS), BigQuery, SQL Server Integration Services (SSIS)

DevOps & Cloud:

Docker, AWS (EC2, RDS, S3), GCP, GitHub Actions, CI/CD, Terraform

Backend & Database Stack:

NestJS, .NET, FastAPI, PostgreSQL, MS SQL Server, Prisma ORM SQL Optimization, Database Performance Tuning

Frontend & UI:

React, Streamlit, Bootstrap

Tools & Methodologies:

Postman (API Testing), VS Code, Jira, Power BI Agile-Scrum, System Design, API Design, Software Testing, Real-time Systems

Employment History

AI Intern / Backend Contributor, Hubinit (Fuuss), Amsterdam, Netherlands - Remote

AUGUST 2024 — MAY 2025

- Built and refactored core backend services using NestJS, TypeScript, PostgreSQL, and Prisma ORM, focusing on modules like AuthService, CardService, and UserService to ensure system stability and scalability.
- Took ownership of early product definition, mapping the full data model, conducting research, and presenting MVP specifications that shaped the project direction in collaboration with the CEO and cross-functional teams.
- Contributed to implementing and refining role-based access control (RBAC) for multi-role dashboards (Super Admin, Business Owner, Customer).
- Collaborated with DevOps on deploying Dockerized services and managing infrastructure using AWS EC2, RDS, S3, CloudWatch, and Amazon SES, while contributing to CI/CD scripting and deployment workflows.
- Investigated and helped mitigate a bot attack, analyzing AWS RDS PostgreSQL logs, implementing dynamic rate limiting, and coordinating reCAPTCHA integration with frontend and management.

- Facilitated seamless frontend-backend integration for the React + TypeScript frontend, defining API contracts and supporting features like wallet and loyalty card management.
- Maintained and extended automated Postman API test suites, ensuring route stability and reducing regressions.
- Collaborated daily in a Scrum-based Agile team, and directly advised the CEO and HR on developer morale, team retention, and internal challenges during organizational transitions.

AI Intern, RoshAI, Kerala, India

MARCH 2025 — MAY 2025

- Contribute to the development of computer vision models for autonomous driving and ADAS, focusing on image-based object detection tasks.
- Curate, label, and refine image datasets using Azure Custom Vision to support training pipelines and improve detection accuracy.
- Customized YOLOv7 for detecting vehicle lights (high/low beam) and pedestrian crosswalks by implementing early stopping, anchor box tuning, image-quality-based sample weighting, and advanced hyperparameter search.
- Integrated additional training hooks into YOLOv7 to enhance convergence and training stability across varied lighting conditions.
- Processed Tata FleetEdge telemetry data for driver behavior analysis, engineered features for smoothness scoring, and built an API-ready FastAPI pipeline for real-time insights.
- Collaborate with cross-functional AI and data teams to iterate on models and support deployment-readiness for production environments.

Data Science Intern, Irohub Infotech, Kerala, India

JUNE 2024 — JANUARY 2025

- Gained experience in NLP, machine learning (regression, clustering), Trees, deep learning (ANN, CNN, RNN, bi-LSTM), and integrated solutions with OpenCV.
- Contributed to real-world projects and delivered results through regular assessments and hands-on tasks.
- Gained hands on experience with tools and platforms such as Streamlit, Excel, Power BI, and SQL.
- Developed data manipulation and visualization skills using NumPy, Pandas, Matplotlib, and Seaborn.

Junior Software Engineer, Permanent TSB, Dublin, Ireland

JULY 2022 — JANUARY 2023

- Took over a critical loan assessment application project during its testing and deployment phase, supporting the migration of a legacy Excel/Visual Basic solution to a modern C#/.NET web application with a Bootstrap frontend.
- Led resolution of underwriter-reported issues by triaging bugs, tracing logic flow across C# controllers, frontend input handlers, and SQL interactions, and delivering fixes for functional and UI-related defects.
- Used Jira as a central coordination tool to log, track, and document all bug resolutions, code locations, logic changes, and status updates—streamlining feedback loops between testers, analysts, and technical leads.
- Worked closely with underwriters to understand business use cases tied to Standard Financial Statement (SFS) data, ensuring accurate handling of treatment workflows, budgeting recommendations, and supervisor reviews.
- Engaged with a large-scale SQL Server backend—analyzing how stored procedures, functions, and data pipelines (owned by the business team) influenced application logic and user experience.
- Delivered the application's initial release milestone by proactively addressing a wide range of bugs and feature requests, despite limited documentation and abrupt developer handoffs.

Junior Database Developer, Datafloat Technologies India Pvt. Ltd., Kerala, India

JUNE 2019 — AUGUST 2020

- Developed and maintained SQL Server databases for leading U.S.-based real estate firms, while gaining experience with Windows Server environments.
- Performed ETL operations across staging and production servers using RETS/RESO APIs, PowerShell, batch scripts, and stored procedures.
- Automated routine database tasks with SQL Server Agent and Windows Task Scheduler, improving process efficiency.
- Collaborated with front-end and metadata teams to meet project deadlines and enhance delivery outcomes.
- Created routine reports using SSRS and XML scripting; conducted data migration across servers using SSIS.
- Used Jira to manage tickets and provide timely resolutions for database-related issues.
- Optimized SQL queries and handled debugging, performance tuning, and deadlock resolution.
- Mentored and onboarded new joiners, helping them understand tools, workflows, and company practices.

Education

Master of Science: Big Data Analytics and Artificial Intelligence, Letterkenny Institute of Technology (Now Atlantic Technological University - ATU), Letterkenny, Donegal, Ireland

SEPTEMBER 2020 — OCTOBER 2021

- Completed an intensive program focused on examining, interpreting, and leveraging large datasets, and enabling intelligent decision-making systems.
- Gained expertise in advanced data analytics, machine learning, and artificial intelligence, with an emphasis on practical application and real-world projects.
- Core Modules: Natural Language Processing, Computer Vision, Data Pre-Processing, Machine Learning, Deep Learning, Hadoop, Apache Spark, Amazon AWS, Databricks, Linux Console, Kafka Confluent, and KSQL.

Bachelor of Engineering: Computer Science and Engineering, Loyola Institute of Technology and Science - Anna University, Nagercoil, Tamil Nadu, India

AUGUST 2013 — MAY 2017

- Acquired foundational knowledge in computer science and engineering principles, focusing on the design, development, and analysis of computing solutions.
- Developed strong problem-solving and analytical skills, applying software engineering methodologies for robust system development.
- Prepared to adapt to emerging information and communication technologies and innovate solutions for complex challenges.

Projects

AI Assistant Hub (Multi-Agent System)

JUNE 2025

- Architected and developed a modular multi-agent AI system using LangGraph to dynamically orchestrate 6+ specialized agents (e.g., Real Estate, RAG-based QA, Summarizer, Translation), enabling diverse and complex query handling.
- Engineered advanced prompt engineering strategies to simulate sophisticated function chaining and reasoning with Google Gemma API and other open-source LLMs, effectively overcoming limitations in native tool-calling.
- Integrated a Retrieval-Augmented Generation (RAG) system with ChromaDB and external APIs (e.g., DuckDuckGo Search) for context-aware information retrieval and real-time data access.

- Designed and implemented an intuitive Streamlit UI featuring chat history, file uploads, and transparent reasoning logs, significantly enhancing user experience and system interpretability.
- Demonstrated proficiency in LangChain and LangGraph by building a robust, extensible agentic architecture entirely with free and open-source resources.

GitHub | Render

Medical AI Chatbot with BioBERT-Transformer Architecture

JUNE 2025

- Developed a medical Q&A chatbot using a custom BioBERT-based encoder and Transformer decoder architecture, optimized for clinical language understanding.
- Trained on 260K+ real-world doctor–patient conversations using TFRecord-like caching, AMP training, label smoothing, and OneCycleLR scheduling in PyTorch.
- Implemented a modular pipeline with early stopping, gradient accumulation, checkpoint recovery, and sample logging for transparent debugging and tuning.
- Built dual Streamlit interfaces using Beam Search and Nucleus Sampling to support controlled and diverse generation modes.

GitHub

MSc Dissertation - A novel racism detection model aiming to minimize bias

AUGUST 2021

- Designed and implemented multi-modal architecture.
- Built a network based on BERT Embeddings and multi-branched CNNs on top of bi-LSTM and a final classification layer for text classification.
- Used a dense neural network for demographic and contextual classification (numerical) completing the architecture.

Github

Heart Attack Prediction in Indonesia - Capstone Project | Data Engineering Zoomcamp

MARCH 2025

- Designed and deployed a scalable data pipeline on GCP using Terraform, GCS, BigQuery, dbt, and Apache Airflow.
- Automated ingestion, transformation, and visualization of heart attack data.
- Built Power BI dashboards and CI/CD workflows with GitHub Actions for real-time insights.

Project GitHub | Course GitHub

Electric Vehicle Market Analysis & Predictive Modeling

DECEMBER 2024

- Analyzed EV market trends and built predictive models for future characteristics through 2030.
- Cleaned and engineered features using regex, imputation, and correlation analysis.
- Built a multi-output regression model in TensorFlow to predict EV price and range.
- Tuned hyperparameters via Bayesian optimization with Keras Tuner.
- Applied DBSCAN clustering with full preprocessing pipelines to identify market segments.
- Visualized findings with time series, heatmaps, and 3D cluster plots using Matplotlib and Seaborn.

GitHub

Fake News Detection System

DECEMBER 2024

- Built an end-to-end fake news classifier using NLP and optimized ML models.
- Engineered text preprocessing with NLTK (tokenization, lemmatization, custom cleaning).
- Tuned TF-IDF parameters and applied Bayesian optimization on XGBoost and Neural Nets.
- Visualized key patterns with word clouds and distribution plots.
- Deployed an interactive real-time Streamlit app for prediction.

GitHub

Vehicle Tracker and Counter using YOLOv8 + Deep SORT

FEBRUARY 2025

- Built a real-time vehicle tracking and counting system using YOLOv8 and Deep SORT.
- Used OpenCV for video processing and interactive virtual line logic.
- Enabled consistent ID tracking across frames and counted unique vehicles across zones.
- Processed real street surveillance datasets for traffic monitoring use cases.

GitHub

Certificates

Data Engineering Zoomcamp, DataTalksClub

APRIL 2025

dbt Fundamentals, dbt Labs

FEBRUARY 2025

Web Design for Everybody: Basics of Web Development & Coding Specialization, University of Michigan (Coursera)

SEPTEMBER 2023

Interests

- College soccer team player and self-taught guitar hobbyist.
- For my non-technical background and skills, including current and interim activities, please visit my Indeed profile: Indeed

Consent

I consent to the processing and storage of my personal data for recruitment purposes in accordance with applicable data protection laws, including consideration for current and future employment opportunities.