# Purushothaman Natarajan

Cuddalore, TN, India | purushothamanprt@gmail.com | +91-9442218587
linkedin.com/in/purushothamann | github.com/Purushothaman-natarajan

## Summary

Applied AI / GenAI Systems Engineer and ML Researcher with 4+ years of industry and research experience building production-grade GenAI, Agentic AI, Computer Vision, and OCR systems across enterprise and safety-critical domains. Specialized in Large Language Models, RAG pipelines, multi-agent automation, multimodal AI, and explainable ML. Proven track record delivering scalable AI platforms that reduce manual effort, improve model reliability, and automate end-to-end workflows across vulnerability intelligence, document AI, and visual understanding systems at Blackstraw AI, Amazon, DRDO-sponsored research programs and with multiple peer-reviewed publications. Strong focus on translating research-driven models into deployable, monitored, and high-impact production systems.

## Skills

**Core AI Specialization:** Generative AI, LLM Systems, Agentic AI, Retrieval-Augmented Generation (RAG), Computer Vision, Multimodal AI, Production ML Systems
**Programming Languages:** Python, SQL, C++ (Basics), JavaScript (Basics)
**LLMs & GenAI:** GPT, Claude, LLaMA, Mistral, Qwen, Prompt Engineering, Fine-tuning, LoRA, Instruction Tuning, Guardrails, Tool Calling
**Agent Frameworks:** DSPy, LangChain, LangGraph, AutoGen, CrewAI, OpenAI SDK, Multi-Agent Orchestration, A2A Communication
**Machine Learning & Deep Learning:** Transformers, CNNs, RNNs, LSTM, Diffusion Models, GANs, Autoencoders, Reinforcement Learning (PPO, DPO), Representation Learning
**Computer Vision & Document AI:** YOLO, SAM, SegFormer, DeepLabV3+, ViT-STR, MGP-STR, Image Classification, Object Detection, Image Segmentation, OCR Systems
**Retrieval & Vector Search:** RAG Pipelines, Semantic Search, Embedding Systems, FAISS, Pinecone, Milvus, USearch
**Natural Language Processing:** Named Entity Recognition, Information Retrieval, Text Classification, Topic Modeling, Privacy & PII Detection
**Data Science & Experimentation:** Statistical Analysis, Hypothesis Testing, A/B Testing, Experiment Design, Metrics Monitoring
**Cloud & MLOps:** Azure, GCP, Docker, Kubernetes, FastAPI, CI/CD, Model Deployment
**Databases:** MySQL, PostgreSQL, Neo4j, LMDB
**Security & AI Risk:** CVE Analysis, Threat Intelligence Ingestion, Risk Scoring, Secure ML Deployment

## Experience

**Data Scientist (Computer Vision & NLP)**, Blackstraw AI – Chennai, IN — Nov 2024 – Present

- Architected and deployed production-grade LLM and agentic AI systems automating enterprise workflows across document intelligence, vulnerability research, and product data pipelines, reducing manual processing effort by up to 91%.
- Led client demos and technical walkthroughs of AI systems, gathered functional and data requirements, and translated business needs into deployable AI pipeline designs.
- Built search-augmented LLM frameworks integrating SERP and external knowledge sources to enable structured reasoning and automated attribute extraction from unstructured technical data.
- Designed scalable computer vision and embedding-based pipelines supporting large-scale retail and logistics image classification and similarity search workloads.
- Engineered vector-search and embedding anomaly detection pipelines improving edge-case classification robustness and similarity matching accuracy.
- Developed advanced OCR and document intelligence stack (SegFormer, DeepLabV3+, ViT-STR, super-resolution models), reducing manual document review effort by 85% while improving downstream text extraction accuracy.

- Built layout-aware document understanding pipelines improving structured field extraction and reading-order consistency in complex enterprise documents.
- Implemented automated PII detection and redaction pipelines across documents and images using NLP + embedding similarity matching.
- Designed multi-agent CrewAI workflows for automated codebase migration (Oracle→MySQL, Java→Python) with validation and tool orchestration stages.

**Research Fellow (Machine Learning & eXplainable AI)**, SRMIST – Chennai, IN Sept 2023 – Oct 2024

- Selected as DRDO-funded PhD research fellow focused on explainable and reliable AI systems for defense-grade sonar imaging applications.
- Developed deep learning models for underwater sonar object detection and classification under low-signal and noisy sensing conditions.
- Built explainability-first ML pipelines integrating LIME, SHAP, and Grad-CAM to produce interpretable predictions for mission-critical AI systems.
- Delivered self-explainable AI models validated and tested by NPOL and DRDO teams for operational underwater surveillance scenarios.

**Machine Learning Associate**, Amazon – Chennai, IN Jul 2022 – May 2023

- Improved annotation and data quality pipelines for Alexa, Ring, and Halo ML systems using NLP preprocessing and normalization workflows.
- Standardized labeling and validation processes, reducing downstream model noise and improving training dataset consistency.
- Built performance dashboards and QA metrics tracking that enabled faster feedback cycles and process optimization.

**Customer Support Executive**, Amazon – Coimbatore, IN Aug 2021 – Nov 2021

- Managed customer and seller queries, refunds, and returns, achieving top performer status for two consecutive months.

## Education

**SRM University**, PhD in Computer Science Feb 2024 – Nov 2024

- **Coursework:** eXplainable AI, Artificial Intelligence, Comparison of Learning Algorithms, and Computational Theory.

**BITS, Pilani**, M.Tech in Data Science Sept 2022 – Sept 2024

- **Coursework:** Data Science, Applied Machine Learning, Deep Learning, Natural Language Processing, Information Retrieval, Artificial and Computational Intelligence.

**Anna University**, B.E in Mechanical Engineering Aug 2015 – Nov 2020

- **Coursework:** Kinematics of Machinery, Thermodynamics, Manufacturing Technology, Internal Combustion Engines, Design and Development.

## Publications

**Synth-SONAR: Sonar Image Synthesis with Enhanced Diversity and Realism via Dual Diffusion Models and GPT Prompting** Sept 2024

***Purushothaman Natarajan***, Kamal Basha, Athira Nambiar

**VALE: A Multimodal Visual and Language Explanation Framework for Image Classifiers using eXplainable AI and Language Models** June 2024

***Purushothaman Natarajan***, Athira Nambiar

**Underwater SONAR Image Classification and Analysis using LIME-based Explainable Artificial Intelligence** Dec 2023

***Purushothaman Natarajan***, Athira Nambiar

## Projects

**Autonomous Vulnerability Intelligence Platform** Enterprise

- Built an autonomous vulnerability intelligence platform using LLM agents to ingest, analyze, and reason over continuously updated CVE and security intelligence sources.
- Implemented agent-based reasoning workflows for vulnerability triage, exploitability analysis, and automated remediation recommendation generation.
- Designed multi-agent orchestration flows for evidence aggregation, risk scoring, and structured remediation packaging.
- Integrated RAG and tool-augmented reasoning agents for continuously updated vulnerability intelligence.
- Tools: Python, LLM Agents, AI Foundry, RAG, Vector DB, Tool Calling, Workflow Orchestration.

**Agentic Product Packaging Intelligence System** Enterprise

- Developed search-augmented LLM classification pipelines combining SERP retrieval and DSPy reasoning to automate standards-based product packaging code mapping.
- Automated product attribute extraction and standards-based packaging code mapping using retrieval + model reasoning.
- Reduced manual lookup and enrichment workload through automated structured extraction pipelines.
- Tools: DSPy, Python, SERP API, LLM Reasoning, NLP Pipelines.

**Retail Vision Intelligence - Detection & Large-Scale Classification on Edge Devices** Enterprise

- Developed edge-deployable computer vision models for retail shelf monitoring and product detection across complex in-store environments.
- Built large-scale product classification pipeline supporting **200K+ product labels** using a custom vector search based classification system.
- Implemented embedding-based retrieval to match shelf products against large product catalogs with improved scalability compared to traditional softmax classification.
- Optimized inference pipelines for mobile and edge devices enabling real-time shelf analytics.
- Tools: YOLO, Image Embeddings, Vector Search, OpenCV, PyTorch, Edge AI Optimization.

**Logistics Asset Detection & Cross-Location Package Similarity Tracking** Enterprise

- Built computer vision pipelines to detect logistics assets including packages, cardboard boxes, pallets, and forklifts across warehouse environments.
- Implemented embedding-based similarity matching to determine whether packages observed across different warehouse locations correspond to the same physical item.
- Improved clustering quality and reduced misclassification using SimCLR-based representation learning and anomaly detection methods.
- Tools: YOLO, SimCLR, CNNs, Embedding Similarity Search, OpenCV, PyTorch.

**Enterprise Document Intelligence Platform - OCR Receipts** Enterprise

- Built a production document AI pipeline converting large-scale receipt and retail documents into structured analytical datasets.
- Improved OCR accuracy using FFT-former deblurring and SwinIR/BSRGAN super-resolution models.
- Reduced super-resolution latency from 16s to 2–3s and improved downstream text extraction quality.
- Tools: OCR, Deep Learning, SwinIR, BSRGAN, FFT-former, NLP, Data Post-processing.

**Agentic Item Description Expansion using LangChain & LangGraph** Enterprise

- Built a multi-agent pipeline to expand abbreviated product titles into rich, structured item descriptions for large-scale e-commerce catalogs.
- Designed LangGraph-based multi-step workflows integrating retrieval, reasoning, and intelligent generation.
- Implemented LangChain tool-calling agents for product metadata retrieval, attribute extraction, and validation.

- Tools: LangChain, LangGraph, LLM Agents, lmdb, Python, Prompt Engineering, Retrieval Pipelines.

**Hierarchical LLM Decision Automation Engine for Item Coding** Enterprise

- Automated playbook generation from manually coded item classification guidelines covering 3600+ different categories, enabling 70% automation of manual item coding workflows.
- Designed hierarchical reasoning pipelines to convert manual classification logic into structured LLM-driven decision workflows.
- Implemented DSPy-based prompt optimization pipelines improving classification consistency and reasoning accuracy.
- Built image-text noise filtering pipelines and rule-based validation layers improving robustness across noisy product metadata inputs.
- Tools: DSPy, Python, Large Language Models, Hierarchical Reasoning, Image-Text Processing, Rule-based Validation.

**Doctane — Modular OCR & Document AI Framework** Open Source

- Built an open-source OCR and document understanding repository integrating state-of-the-art models.
- Implemented modular pipelines for text extraction, preprocessing, and document intelligence tasks.
- Repo: github.com/Purushothaman-natarajan/doctane
- Tools: Python, Transformers, PyTorch, NLTK, Streamlit.

**Explainable Remote Sensing Vision Framework** Research

- Built a transfer learning-based remote sensing scene classifier with model explainability.
- Applied LIME and Grad-CAM to generate visual explanations for model predictions.
- Repo: github.com/Purushothaman-natarajan/eXplainable-AI-for-Image-Classification-on-Remote-Sensing
- Tools: Python, TensorFlow, Scikit-learn, LIME, Grad-CAM, Gradio.

## Additional Experience and Awards

**Business Analyst, IIFL (Oct 2019 – May 2021):** Managed daily client transactions worth $10M, liaised between clients and the trading desk, and provided strategic market insights.

**Instructor, BrightNext Academy (2023-2024):** Taught Machine Learning and Deep Learning courses to over 100 students.

**Freelancer (Upwork and LinkedIn) (2022-2024):** Successfully Delivered AI and ML projects valued between $10K to $20K as an independent contributor.

**Third Prize, Innovation & Design on Remote Sensing Data by Hack2Skill:** Recognized for designing a synchronized research and production-ready dashboard for an explainable image classifier, competing against 100+ entries.

## Licenses & Certifications

**Udacity:** Deep Learning, Computer Vision, Generative AI Nanodegree's.

**DeepLearning.AI:** Retrieval Augmented Generation (RAG), Reliable AI using Guardrails.

**Codecademy:** BI Dashboards with Power BI & Tableau, Business Intelligence Data Analyst, SQL for Marketers and Product Managers.

**LinkedIn:** Advanced SQL, Artificial Intelligence Foundations: ANN, CNN, R-CNN, RNN, LSTM, GNN & Transformers, Advanced NLP with Python for Machine Learning, GANs and Diffusion Models with TensorFlow and PyTorch, Transfer Learning Using PyTorch, Deep Learning for Computer Vision Applications.