

VISHNU SHREERAM M.P.

+91 9633695099 142201014@smail.iitpkd.ac.in linkedin.com/vishnu-m-p github portfolio

Carrer Objective

Aspiring researcher looking to bridge the gap between theoretical guarantees and practical efficiency in AI. With a background spanning Generative Models and Industrial optimization, I aim to join a research team where I can contribute strong engineering skills while investigating the fundamental mathematical challenges of next-generation machine learning.

Education

Indian Institute of Technology, Palakkad Expected May 2026
Bachelor of Technology in Data Science (CGPA: 9.6 / 10) Palakkad, Kerala

- Relevant Courseworks:** DSA, Optimisation, Artificial Intelligence, Database Systems, Computer Systems for Data Science, Discrete Maths, Probability & Statistics, Linear Algebra, Multivariable Calculus, Machine Learning, Data Mining, Data Analytics, Deep Learning, **MLOps**, Big Data Lab, Natural Language Processing, AI Ethics, Information Retrieval, **Gen AI**, Probabilistic Graphical Models, Multi-Agent Systems, Computer Networks.

Experience

Laboratory of Statistical Artificial Intelligence and Machine Learning May 2025 – July 2025
Research Intern Palakkad, Kerala

- Collaborated on developing a hybrid training method for Probabilistic Circuits (PCs), more generalizable than EM and faster than SGD.
- Worked on a Hessian-based regularizer to reduce overfitting by guiding models toward flatter optima.
- Demonstrated that the Hessian trace, a sharpness proxy, is efficiently computable for PCs and enables closed-form updates.
- Trained neural networks and optimized hyperparameters on HPC infrastructure using distributed resources, validating the method across multiple probabilistic circuit benchmarks.
- Co-authored - "Tractable Sharpness-Aware Learning of Probabilistic Circuits" which was accepted for Oral Presentation at AAAI-26 (Association for the Advancement of Artificial Intelligence), Singapore, 2026. and preprint available on [arXiv](#).

Laboratory of Statistical Artificial Intelligence and Machine Learning May 2024 – April 2025
Part-time Research Intern Palakkad, Kerala

- Worked on optimizing Bosch's SMT assembly line for PCB production, involving 32 PCB types and 200+ unique components.
- Developed an algorithm to determine optimal feeder arrangements for each product group, improving pick-and-place efficiency under real-world constraints.
- Improved components-per-hour (CPH) by at least 7% while meeting Bosch's practical manufacturing constraints.
- Patent filing underway due to the novelty of the approach.
- Applied Genetic Algorithms, Mixed-Integer Linear Programming (MILP), and Nonlinear Programming techniques and worked with OR-Tools and Gurobi optimizer

Projects

Viśva Mitra – Voice Enabled Agentic AI Assistant | Python, LangChain, Ollama, MCP, Docker, FastAPI, uv | Github Repo

- Built an AI assistant that handles both chat and voice queries, using Whisper for speech-to-text and open-source TTS models for spoken responses, and intelligently invokes tools to generate responses.
- Supports 15+ actions including computer control, real-time weather, web search, and more.
- Containerized MCP-based tool servers using Docker and implemented tool routing via LangChain and Ollama.
- Designed a modular SaaS architecture with clean service separation, robust error handling, result tracking via MLflow, and workflow management using Prefect. Ensured maintainability with uv, Ruff, Justfile, and TOML/YAML.
- Implemented workflow orchestration and CI/CD-like deployment pipelines via Prefect & Docker.

ViZearch – Semantic Search over Visual Data | Python, FastAPI, BentoML, Tantivy, Locust | Github Repo

- Built a semantic search engine for images & videos using BLIP for captioning and Tantivy for indexing & querying.
- Deployed dual backends with FastAPI and BentoML and load tested for over 1000 requests using Locust.
- Ensured clean, configurable microservices with unified logging, ruff, and TOML/YAML config management.

Central Instrumentation Facility (CIF) Database Management System | PostgreSQL, Python | Github Repo

- Developed a comprehensive Slot Booking Database Management System using PostgreSQL.
- Incorporated 25 necessary triggers, functions & procedures adhering to the requirements of CIF at IIT Palakkad
- Collaborated in a team of 4 to build an intuitive front-end & back-end supporting role-based access for students, faculty & staff.

Technical Skills

Languages: Python, PostgreSQL and familiarity with C++, Java, MATLAB, R, Scala

Libraries and Frameworks: TensorFlow, PyTorch, Scikit-learn, Numpy, Pandas, ReactJS

Technologies: Model Context Protocol, Docker, Apache Spark, Hadoop & Hive, FastAPI, Git, Linux, Jupyter, Vector DBs (LanceDB & Pinecone), Amazon Web Services

Awards and Achievements

Certificate of Academic Excellence – Certificates awarded by IIT Palakkad for securing the **highest CGPA** among **first-year** (2022-'23) and **second-year** (2023-'24) students in Data Science Department.

GATE 2025 – Secured **All India Rank - 73** in Data Science and Artificial Intelligence.

Inter IIT Tech Meet 13.0 – Represented IIT Palakkad at IIT Bombay for Pathway's problem statement on *Dynamic Agentic Retrieval-augmented generation (RAG)* for faster and efficient Information Retrieval; secured **12th place out of 23 IITs**.

Patents and Publications

Accepted Conference Paper: Hrithik Suresh, Sahil Sidheekh, Vishnu Shreeram M. P., Sriraam Natarajan, Narayanan C. Krishnan. "*Tractable Sharpness-Aware Learning of Probabilistic Circuits*". Accepted at **AAAI-26** (Association for the Advancement of Artificial Intelligence), Singapore, 2026.

Patent Filed: **Title:** METHOD, APPARATUS, AND SYSTEM FOR INCREASING THROUGHPUT IN A SURFACE MOUNT TECHNOLOGY ASSEMBLY LINE.

Patent Application No: 202541068350.

Filed on: July 17, 2025.