



✉ pranavprakash963@gmail.com

🔗 [Pranav's Pocket](#)

📞 +91 7358589758

🌐 [Linkedin](#)

🐙 [GitHub](#)

Core Skills

- ▷ RAG Pipeline Design
- ▷ Agentic AI Workflows
- ▷ LLM Integration (Claude, Polly, Transcribe)
- ▷ Prompt Engineering
- ▷ Hybrid Search Systems

Frameworks

- ▷ PyTorch
- ▷ TensorFlow | Keras
- ▷ LangChain | LangGraph
- ▷ Scikit-Learn | OpenCV
- ▷ NextJS | TailwindCSS

Languages

- ▷ Python
- ▷ C | C++
- ▷ Java | JavaScript
- ▷ MATLAB

Pranav Prakash Chandra

AI/ML Engineer with production experience building RAG pipelines, Agentic AI workflows, and hybrid-search chatbots using Claude 3, and AWS speech services, with a research background in deep learning for medical image analysis bringing an experimental approach to engineering. Proven ability to own complex AI features end-to-end — spanning system architecture, REST API design, database schema, and prompt engineering.

Work Experience

Software Engineer (GenAI)

Dec 2024 - Feb 2026

Servion Global Solutions Pvt Ltd, India

- Architected and deployed hybrid-search AI chatbots combining vector retrieval and keyword search, improving response relevance and contextual accuracy
- Designed production-grade RAG pipelines integrating Claude 3 with speech-to-speech models (Nova Sonic, Polly, Transcribe)
- Designed and deployed a scalable agentic AI orchestration workflow leveraging Claude and MongoDB-backed state management, productionized as a timed systemd service on Amazon Linux.
- Developed REST APIs and optimized database schemas to support high-throughput AI workloads
- Led prompt engineering and evaluation strategies to enhance model reasoning, response consistency, and task reliability

Machine Learning Engineer

Jul 2024 - Nov 2024

Homy, Sweden

- Developed stateful, multi-turn RAG pipelines using GPT-4, LangChain, LangGraph, and Pinecone, enabling document-grounded QA with persistent conversation memory.
- Designed and deployed FastAPI services to production, including endpoint versioning, request validation, and async inference handling.
- Prototyped multi-agent orchestration frameworks to evaluate reasoning chains and memory persistence, informing the long-term agentic architecture roadmap.

Research Experience

Independent Researcher

Jan 2024 - Jul 2024

Under supervision of Dimtrios Toumpanakis (M.D)

- Developed and evaluated ensemble neural network architectures to improve prediction accuracy and robustness of BrainAge estimation.
- Applied Explainable AI techniques, including Grad-CAM and saliency maps, to interpret model behavior and identify salient brain regions influencing predictions.
- Leveraged interpretability insights to iteratively refine model architecture and training strategy, improving overall pipeline performance.
- Conducted extended research on optimization strategies for BrainAge prediction, focusing on model generalization and feature relevance.

Database Technologies

- ▷ MySQL | PostgreSQL
- ▷ MongoDB
- ▷ Pinecone
- ▷ OpenSearch
- ▷ Neo4j

Education

Master of Science in Image Analysis and Machine Learning

Aug 2021 - Dec 2023

Uppsala Universitet

Thesis Project: Comprehensive Study of Brain Age Prediction using Classical Machine Learning and Neural Networks

BE in Computer Science and Engineering

Jun 2016 - Jul 2020

Sri Sai Ram Engineering College

CGPA: 7.56

Final Project: Monitoring Mental Health using Physiological Signals

Projects

Understanding and Optimizing BrainAge prediction

(12/2023 – 06/2024)

- Implementing Ensemble Neural Network Models to improve and optimize the BrainAge prediction pipeline.
- Implementing gradCAM, Saliency Maps and other XAI algorithms on a custom convolutional neural network to understand its training process for the brain age prediction and using the insights to improve the prediction pipeline.

Prediction of Brain Age using Conventional Machine Learning and Neural Networks

(02/2023 – 12/2023)

Implementing executable brain age prediction pipeline with Machine learning and Deep learning techniques using ScikitLearn and PyTorch.

Comparison of Deep Learning and Machine Learning Models in Classification of Hand drawn Spiral for Parkinson's Diagnosis

(11/2022 - 01/2023)

Provide a comparative analysis of the performance of several machine learning models in identifying Parkinson's and control patients from the features extracted from a pre-trained CNN model (ResNet50)

Vertebrae Segmentation and Labelling for Radiotherapy Application

(09/2022 - 01/2023)

Develop software to segment and label the vertebral bodies in CT images. Integrate the method into RayStation using its scripting API which is written in Python and Visualize the results in RayStation.

Monitoring Mental Health using Physiological Signals

(10/2019 - 03/2020)

Monitoring the emotional state and overall emotional health at regular intervals every day by maintaining a record of the patient's emotional health data.