DEEPANSHU KANDPAL

Senior Software Engineer, AI/Machine Learning
Specializing in Large Language Models & Autonomous AI Agents

Email: dkandpalz@gmail.com   **Phone:** +91-8087669991   **Location:** Mumbai, Maharashtra   **Portfolio:** https://deepskandpal.github.iodeepskandpal.github.io
LinkedIn: https://linkedin.com/in/deepanshu-kandpallinkedin.com/in/deepanshu-kandpal

PROFESSIONAL SUMMARY

Senior Software Engineer with 8+ years of experience designing, building, and deploying production AI systems and autonomous agents at scale. Expert in leveraging Large Language Models (LLMs) and deep learning techniques to solve complex software engineering challenges, from architecture design to deployment. Proven technical leader with experience in large-scale system design, cross-functional collaboration, and end-to-end ML solutions. Creator of **LangChef**, an open-source platform accelerating Generative AI development workflows.

CORE COMPETENCIES

Software Engineering & AI/ML

Large Language Models (LLMs)   Autonomous AI Agents

Deep Learning & Neural Networks

Natural Language Processing   LangChain   Transformers

Hugging Face   Vector Databases (Qdrant)   RAG Systems

Multi-Agent Systems (MAS)   LLM Post-training

System Design & Architecture

Distributed Systems   Microservices

Event-Driven Architecture   System Scaling (100K+ QPS)

High-Throughput ML Pipelines

Programming & Data

Python   Swift   SQL   Metal (GPU)

Data Structures & Algorithms   PySpark   BigQuery

Databricks

Cloud & DevOps

AWS (Bedrock, SageMaker, Lambda)   Azure   GCP

Kubernetes   Docker   Terraform   MLOps   MLflow

Redis   CI/CD

PROFESSIONAL EXPERIENCE

Senior Machine Learning Engineer
KnowBe4

📅 2024 – Present   📍 Bangalore, India

- Architected and deployed **production GenAI applications** using Claude Sonnet 3/3.5 and AWS Bedrock, including automated phishing email generation system that reduced admin effort by 80% through advanced prompt engineering and input validation workflows.

- Built **RAG-based policy compliance system** that automatically generates quizzes from PDF documents with custom chunking algorithms supporting multiple languages, solving context window overflow challenges for diverse language scripts.

- Developed **multi-label text classification pipeline** using Claude Sonnet 3.5 for Risk Score mapping, automatically categorizing policy content into security buckets (Email, Data, Web, Account Hygiene, Compliance, Physical) with 80%+ confidence thresholds.

- Designed and implemented **advanced recommendation algorithm** using Normal Distribution Sampling to replace hard top-k random selection, providing smoother exploration-exploitation balance for training module recommendations with configurable Standard Deviation Coefficient.

- Led technical architecture for **serverless ML infrastructure** using Terraform, Docker, and AWS Lambda, enabling rapid deployment and scaling of GenAI applications across cybersecurity awareness verticals.

Machine Learning Engineer, Experimentation
FarFetch

📅 2022 – 2024   📍 Gurugram, India

- Architected and maintained **high-throughput ML experimentation platform** processing **100,000+ queries/day** on **300TB+** datasets, directly informing 20% of key business decisions across multiple product teams.

- Led **system migration** of monolithic ML services to **Azure Kubernetes Service (AKS)** with Helm charts, implementing horizontal scaling, reducing deployment lifecycle by 50%.

- Designed and implemented **Multi-Armed Bandit (MAB)** algorithms including UCB and Thompson Sampling for dynamic experiment allocation, and **Mixed Sequential Probability Ratio Testing (MSPRT)** for early stopping mechanisms.
- Solved complex **distributed computing** challenges including Redis-based distributed locking for parallel experiment processing and BigQuery optimization for petabyte-scale data processing.

Data Scientist, Marketing Analytics
Accenture Strategy & Consulting

📅 2021 – 2022   📍 Gurugram, India

- Developed **geospatial ML systems** for outlet prioritization using geospatial features and multi-class store classification, enabling CPG expansion strategies across Southeast Asia markets with 15% improvement in market penetration.
- Architected end-to-end **automation pipelines** that reduced 3-week manual reporting processes to 2-hour automated workflows, processing demographics and econometric data from multiple APIs (Global Data, Euromonitor).
- Solved multilingual store classification challenges (Thai, Vietnamese, Hindi, English) using transfer learning techniques, enabling accurate grocery vs non-grocery classification from mixed-language store names and attributes.
- Delivered market sizing models and quarterly insights that directly influenced CPG strategy decisions for major clients, transforming manual Excel-based processes into scalable Python-driven analytics.

Software Engineer & ML Engineer
Jio Platforms Limited

📅 2017 – 2021   📍 Mumbai, India

- Architected end-to-end **conversational AI systems** for MyJio app (50M+ users), implementing intent classification with LSTM + FastText embeddings and solving multilingual (Hinglish) NLP challenges through transfer learning.
- Developed complex real-time video processing system using **Metal (GPU programming)** for green-screen removal and dynamic background replacement in iOS streaming applications for HelloJio customer service.
- Built production-ready ML training and inference pipelines with Redis-based model versioning system, enabling zero-downtime model updates across horizontally scaled Kubernetes deployments.
- Served multiple roles as iOS developer, backend engineer, and ML engineer, delivering complete software solutions from mobile client to ML inference infrastructure.

PROJECTS

Project EKΛ (Eka)

Large-Scale Indic LLM Development

📅 2024 – Present   📍 Soket AI: https://eka.soket.ai/eka.soket.ai

- Contributing to development of **100+ billion parameter foundation models** and instruction-tuned models specifically designed for India's linguistic diversity, working with 2T+ token training corpora across multiple Indic languages.
- Collaborating on **energy-efficient training algorithms** and high-quality data practices for culturally representative datasets, ensuring ethical AI development for India's diverse linguistic landscape.
- Participating in **open-source model development** with transparent "build in public" approach, contributing to India-specific benchmarks and evaluation frameworks for multilingual LLM assessment.

LangChef

Open-Source Platform for Generative AI Experimentation

📅 2024 – Present   📍 GitHub: https://github.com/deepskandpal/langchefgithub.com/deep

- Architected and developed **LangChef**, an open-source framework accelerating LLM prototyping and evaluation, enabling rapid experimentation with GenAI workflows including prompt engineering, RAG integration, and multi-agent orchestration.

EDUCATION

Bachelor of Engineering, Electronics and Telecommunication (First Class)

Savitribai Phule Pune University, MIT Academy of Engineering

📅 2013 – 2017   📍 Pune, India