

PRAJYOT KORE

Senior AI Engineer | Machine Learning Researcher

Phone: +91-9579545875 | Email: pakore1999@gmail.com | Portfolio: prajyot.me | LinkedIn: linkedin.com/in/prajyot-kore | GitHub: github.com/PrajyotKore

Senior AI Engineer with expertise in **Agentic AI**, **Multi-Agent Systems**, **Reinforcement Learning**, and **LLM Post-Training**, building enterprise-scale autonomous AI systems deployed across BFSI and Legal AI. Experienced in architecting production-grade AI platforms using, **MCP**, **A2A communication**, **RLHF**, and distributed agent orchestration with a strong foundation in statistical learning from IIT Bombay.

EDUCATION

Degree	Specialization	Institute	Year	GPA
M.Sc.	Applied Statistics & Informatics	IIT Bombay	2022–2024	7.03/10
B.Sc.	Statistics	Shivaji University	2018–2021	9.18/10

PROFESSIONAL EXPERIENCE

ANAPTYSS

Nov 2024 – Present

AI Lead - Multi-Agent Systems | Enterprise Agentic AI (Legal Tech & BFSI)

• ANA — Control Testing Simplified

anaptyss.com/ana

Agentic orchestration. Hybrid supervisor/hierarchical conversational AI automating BFSI control testing across 7 steps with actor-critic validation.

- Architected a **FastMCP**-based hierarchical multi-agent platform with **Model Context Protocol (MCP)** and **Agent-to-Agent (A2A)** communication, enabling scalable orchestration of **17+** specialized agents for enterprise control testing
- Designed **context engineering** and hybrid **memory architecture** (session, episodic, semantic) to support long-running agent workflows, adaptive planning, and cross-agent knowledge sharing across complex audit tasks
- Engineered resilient long-running workflows with checkpointing, recovery mechanisms, actor-critic validation, and human-in-the-loop governance, improving reliability for enterprise-scale autonomous execution

• CovenAce — Covenant Extraction at Scale

anaptyss.com/covenace

Document intelligence. Self-correcting extraction pipeline for financial covenants from legal agreements at production scale.

- Engineered self-correcting validation system with parallel agent architecture, achieving **94.47% accuracy** and **0.93 recall** in entity classification while maintaining **98% consistency** across document types
- Optimized batch inference pipeline with caching strategies, achieving **80–85% faster processing** and up to **70% cost reduction**, enabling real-time covenant monitoring for financial institutions

• Index AI — Multi-Agent Document Indexing

anaptyss.com/index-ai

Ensemble document indexing. OCR on combined-scan banking documents resolved by per-page majority voting across parallel agents.

- Built ensemble multi-agent document indexing system combining OCR on combined-scan banking documents with N parallel classification and segmentation agents per page
- Resolved per-page classification via majority voting, achieving **98% accuracy** on synthetic silver dataset across diverse banking document types

RIPIK.AI

Jun 2024 – Oct 2024

ML Engineer – Computer Vision & Optimization | Industrial AI Solutions

- Compressed **Segment Anything Model (SAM)** for edge deployment using **ONNX quantization** and structured pruning, reducing VRAM by **43%** and inference latency by **32%** while maintaining **95% accuracy**
- Deployed production **YOLOv8** fire detection system on AWS EC2 with real-time video processing, achieving **85% mAP@50** through synthetic data augmentation and temporal consistency filtering
- Built hybrid optimization engine combining Genetic Algorithms and Linear Programming for manufacturing scheduling, increasing client operational efficiency by **18%** and net profit by **5%**
- Implemented **MLOps** pipeline with Docker and AWS for model versioning, A/B testing, and automated deployment

- Developed multimodal late-fusion architecture combining **BERT** (text) and **Wav2Vec2.0** (audio) for inner emotion recognition, achieving **93% accuracy** on Japanese emotional speech dataset
- Implemented ROI-based facial landmark extraction pipeline for video analysis with temporal modeling, achieving **0.91 micro-F1** and reducing false positives by **23%**
- Conducted ablation studies on fusion strategies (early vs. late) and published findings in internal research reports

RESEARCH EXPERTISE

AI/ML & Agentic AI: Multi-Agent Systems, Agentic AI, LLM Agents, Hierarchical Agent Orchestration, Agent Memory, Context Engineering, Long-Running Workflows, Tool Calling, **FastMCP**, **MCP**, **A2A**, **RLHF** (PPO, DPO, GRPO), Meta-Reinforcement Learning, World Models, RAG

LLM Training & Alignment: Supervised Fine-Tuning (SFT), Reward Modeling, Preference Optimization, **LoRA**, **QLoRA**, HuggingFace **TRL**, Synthetic Data Generation, LLM Evaluation

Frameworks & Libraries: **PyTorch**, TensorFlow, JAX, HuggingFace, TRL, PEFT, Gymnasium, LangChain, LangGraph, Google ADK, FastMCP, ONNX

Engineering & Infrastructure: GCP, **Vertex AI**, Kubernetes, Docker, AWS (EC2/S3), Redis, PostgreSQL, ChromaDB, SQL, Git, System Design, MLOps

RESEARCH & KEY PROJECTS

MAP-Boosting: Multi-Agent Prompt Optimization

Under Review

- Proposed novel framework treating prompt optimization as pseudo-gradient descent over discrete instruction space using ensemble of specialized agents with weighted aspect-based feedback
- Designed boosting-inspired failure mode prioritization mechanism to systematically address domain-specific weaknesses in LLM reasoning
- Achieved SOTA performance: **+11.7% on AIME 2024**, **+10.6% on SWE-bench Verified**, **+8.5% on Live-CodeBench**, outperforming DSPy and OPRO baselines
- Developed an automated LLM evaluation framework combining benchmark-driven assessment, synthetic test generation, reward-based optimization, and iterative feedback loops

End-to-End RLHF Pipeline for LLM Alignment

Independent Research

- Built end-to-end LLM post-training pipeline implementing SFT, Reward Modeling, PPO, DPO, and GRPO using **HuggingFace TRL**, with parameter-efficient adaptation via **LoRA** and **QLoRA**
- Trained reward model on Anthropic HH-RLHF dataset using Bradley-Terry loss, achieving **87% agreement** with human preferences and **0.82 rank correlation**
- Optimized policy with PPO + adaptive KL-divergence penalties, improving helpfulness scores by **34%** while maintaining safety

RAG-Based Code Visualization System

Jan 2025

- Built Retrieval-Augmented Generation system with ChromaDB and Gemini 2.0 to generate complex Manim animation code from natural language mathematical descriptions
- Designed hierarchical chunking strategy with semantic overlap to handle long contexts for calculus, linear algebra, and probability theory

Tumor Detection on Whole Slide Images (WSI)

M.Sc. Project, IIT Bombay

- Fine-tuned KimiaNet with DenseNet-121 backbone for multi-organ tumor detection on gigapixel WSIs, achieving **89% AUC** on test set
- Engineered preprocessing pipeline with Reinhard color normalization and adaptive patch extraction across kidney, breast, and colon tissues

SCHOLASTIC ACHIEVEMENTS

- **All India Rank 22** in GATE Statistics 2023 among 5,000+ candidates
- **All India Rank 69** in ISI M.Stat Entrance Examination 2022
- Secured **97.8 percentile** in IIT JAM Statistics 2022