

NITIN KUKREJA

AI Research Engineer | LLM Post-Training, Reinforcement Learning & Evaluation
Mumbai, India | nkukreja25@gmail.com | [ORCID 0009-0004-2358-7817](https://orcid.org/0009-0004-2358-7817)
nitinkukreja0709.github.io | linkedin.com/in/kukrejanitin | github.com/nitinkukreja0709

Research Interests

Post-training of large language models (SFT, preference optimisation, RL with verifiable rewards); RL environments and reward design for tool-using agents; evaluation of LLM safety and reliability in high-stakes, multilingual (Indic) domains.

Publications

- S. Paul, C. Bhushan, H. Sharma, **N. Kukreja**, Y. Dedhia, K. Doshi, P. Devadiga. *IndicBankBench: Evaluating Safety and Reliability of Language Model Assistants in Indian Retail Banking*. [arXiv:2609.29167](https://arxiv.org/abs/2609.29167), 2026. [\[code\]](#)
- NPCI AI Research Team (18 authors, alphabetical; incl. **N. Kukreja**). *FiMI Banking: A Sovereign Model for Indian Retail Banking*. [arXiv:2609.03960](https://arxiv.org/abs/2609.03960), 2026.
- NPCI (40 authors, alphabetical; incl. **N. Kukreja**). *FiMI: A Domain-Specific Language Model for Indian Finance Ecosystem*. [arXiv:2602.05794](https://arxiv.org/abs/2602.05794), 2026.

Research Experience

National Payments Corporation of India (NPCI) – AI Research Team Mumbai, India
Senior Associate Data Scientist (Apr 2025 – Present); Associate Data Scientist (Sep 2023 – Mar 2025) Sep 2023 – Present

RL post-training for safe tool-using bank agents (FiMI Banking)

- Compared preference optimisation (response-level) with verifiable-reward RL (trajectory-level) for multi-turn banking tool use: out-of-scope refusal **52%→80%**, edge-case score 0.509→0.718, order-sensitive 0.590→0.679, with **29% fewer tokens**.
- Designed **Banking Gym**, a deterministic NeMo Gym-compatible RL environment: stateful bank simulator, confirmation-gated write actions, verifier-graded end-state rewards and parallel isolated rollouts.

Evaluation methodology for banking assistants (IndicBankBench)

- Co-built a 799-case, 20-axis benchmark that scores trajectories in four stages (safety, action/tool use, adequacy, advisory quality) with deterministic gates instead of final-answer-only LLM judging; evaluated 11 models from 8 families.

Domain adaptation of a 24B LLM (FiMI)

- Led post-training and agentic integration: multi-stage CPT → IFT → SFT on **68B** curated English/Hindi/Hinglish tokens (from a 30T-token pool via PII anonymisation, deduplication, semantic validation); **+20%** over Mistral Small 24B on finance reasoning, **+87%** on domain tool calling.
- Synthetic data generation for tool-calling and multi-turn SFT; distributed training on **64 H100s** (DeepSpeed ZeRO-3, FlashAttention-2, Slurm).

Deployment and earlier work

- Took FiMI to production as the multi-agent backend of UPI Help / MyUPI (LangGraph, MCP; RAG/Graph-RAG; guardrails, faithfulness checks); launched at GFF 2025.
- GNN fraud models scoring 15,000 transactions/sec at 20 ms; graph-traversal AML cut manual review by 80%; work featured by NVIDIA at SuperComputing '24.

Engineering Experience

Xoriant Solutions Pune, India

Software Engineer (Associate Software Engineer until Aug 2022)

Aug 2021 – Sep 2023

- LLM chatbot (GPT-3.5, LangChain, Azure Cognitive Search) cutting a client's support load by 40%; document AI with LiLT/BERT entity extraction at 95% accuracy.

Methods & Tools

Methods: SFT, CPT, PEFT/LoRA, preference optimisation, RLVR/GRPO, reward & verifier design, synthetic data, LLM-as-judge vs. deterministic evaluation, RAG/Graph-RAG, GNNs

Systems: PyTorch, Hugging Face, DeepSpeed, FlashAttention-2, NeMo Gym, vLLM, Triton, Slurm, Kubernetes, LangGraph, MCP

Education

R. C. Patel Institute of Technology

B.Tech, Computer Science & Engineering – CGPA 8.4/10

Shirpur, India

2017 – 2021