

# Hemanth Kari

AI/ML Engineer · Full-Stack and Frontend Engineer · Python, PyTorch, TypeScript, React, Next.js, LLMs, RAG

[hckariusa@gmail.com](mailto:hckariusa@gmail.com) | +1-312-583-7830 | Open to relocation / remote | [github.com/HemanthKari](https://github.com/HemanthKari) | [hemanthkari.hackthru.com](https://hemanthkari.hackthru.com) | [huggingface.co/HemanthKari](https://huggingface.co/HemanthKari)

Authorized to work in the US. No sponsorship required.

## Education

---

**M.S. Computer Science**, Purdue University Northwest, Aug 2022 – May 2024 (GPA 3.5/4.0). Coursework: machine learning, distributed systems, algorithm design, database systems, computer security.

**B.Tech. Biotechnology**, KL University, India, Jul 2017 – May 2021

## Experience

---

**ASuiTech Solutions**, Full-Stack Software Engineer (Contract), Remote *Dec 2025 – Present*

- Rebuilt the production order-dispatch pipeline (200+ orders/day) as idempotent Kotlin Kafka stages with DynamoDB step state, designed for horizontal scalability. Failures self-heal via Kafka log replay instead of manual recovery scripts.
- Diagnosed a recurring overselling bug in production on high-demand SKUs. The polling-based inventory sync had a race condition. Replaced it with an SQS / Lambda event flow plus Redis locks at the SKU level.
- Moved the product and checkout pages to Next.js 14 server components. Core Web Vitals (LCP) on owned pages improved 12 percent. Separately, debouncing the React catalog filters and trimming the Elasticsearch query cut retrieval latency 30 percent.

**Horizon IT**, Software Engineer, Remote *Jun 2024 – Nov 2025*

- Migrated client state. Server state moved to React Query, UI state to Zustand. Cut redundant refetches by 40 percent. Added retry-with-backoff plus a unified error boundary so transient API failures stopped showing up as broken pages.
- Accessibility pass on the assigned flows: Lighthouse score from 78 to 95 after replacing custom interactive widgets with semantic primitives and fixing keyboard / focus handling. Grew Jest and React Testing Library coverage by 22 percent so regressions caught in CI, not PR review.

**Deepcloud**, Software Engineer, Bangalore, India *Jan 2021 – Aug 2022*

- Wrote a TypeScript provisioning service that automated the full VM lifecycle for dev environments across the engineering org, replacing a brittle 30-minute manual checklist.
- Hardened Python and Spring Boot microservices that had been compounding failures under load: circuit breakers on the failing-most calls, retry-with-backoff on the recoverable ones, graceful-degradation paths where partial data was acceptable. Profiled and debugged the PostgreSQL and MongoDB hot paths, added the missing indexes, rewrote a pair of N+1 query patterns on the worst route. Cut peak latency on the worst routes by 45 percent.
- On-call rotation. Wrote runbooks, remediation scripts, and post-mortems for each recurring incident so the next page resolved with a known command. MTTR went from 30 to 10 minutes.

## Projects

---

**FluxionDock**: retrieval-augmented ML system for protein-ligand pose prediction (in progress; US provisional patent application in preparation) | PyTorch, FAISS, FastAPI, Next.js [[Dataset](#) · DOI: [10.57967/hf/8408](https://doi.org/10.57967/hf/8408) · [Code \(Zenodo\)](#)]

- Retrieval-augmented flow-matching pose-prediction model in PyTorch (2.82M research / 23.2M shipping params). Retrieval keeps the generator small via analogous-complex conditioning; flow matching gives single-pass deterministic inference. Full training: 37 minutes over 10 epochs on one A100.
- Research model: 1.60 Å mean Kabsch RMSD with 69.8 percent of poses below 2 Å on a 3,608-complex held-out validation set. The shipping baseline is evaluated leave-one-out, so retrieval cannot leak into the test conditioning. It reaches 1.49 Å / 75.3 percent.
- Built the retrieval substrate: 72,199 RCSB PDB complexes, ESM-2 pocket embeddings + RDKit Morgan ligand fingerprints, indexed in FAISS HNSW (vector search) for sub-second lookup. CC-BY 4.0 on Hugging Face with reproducible GitHub Actions CI/CD, 900+ downloads. Validated via SIFTS same-target evals: 65.1 percent same-UniProt at rank 1 (median cosine 0.982 vs 0.440 random).

- Diagnosed why MoLFormer-XL (preferred ligand encoder) would not load: its tokenizer files were incompatible with the pinned transformers version. Re-pinning the stack for one component was the wrong v0.1 trade. Fell back to RDKit Morgan, logged the active backend in config.json, kept the transformer path open for v0.2.

**Llama-3.1 Pro Coder v1:** code-specialist LLM fine-tune | PyTorch, QLoRA, Hugging Face [[Hugging Face](#)] · DOI: [10.57967/hf/8407](#)]

- Fine-tuned Meta Llama-3.1-8B-Instruct with QLoRA (rank 128) on 111,289 commercial-license code samples (48M tokens); 4-bit quantization let the 8B model fit in 40GB VRAM on one A100. Diagnosed a format mismatch between verbose CoT training data and HumanEval's zero-shot harness that was depressing scores even when underlying coding competence improved. 68.3 percent pass@1 on HumanEval; 4-bit checkpoint released on Hugging Face, 1,200+ downloads.

**Swale:** Android launcher app on Google Play | Flutter, Firebase, RevenueCat [[Google Play](#)]

- Android launcher shipped on Google Play with RevenueCat subscriptions and Firebase for analytics and crash reporting. 100 downloads.

## Publications

---

ORCID: [0000-0001-5755-2242](#) · [Google Scholar](#)

**Kari, H.**, Bandi, S. M. S., Kumar, A., Yella, V. R. "DeePromClass: Delineator for Eukaryotic Core Promoters Employing Deep Neural Networks." *IEEE/ACM Transactions on Computational Biology and Bioinformatics* 20(1):802–807, 2023. doi:[10.1109/TCBB.2022.3163418](#). First-author CNN-LSTM, 93.6 percent test accuracy on *C. elegans*; 11 Google Scholar citations (8 received in 2025), cited in *Trends in Biotechnology* (Cell Press, 2024 IF 14.3).

**Kari, H.** "Pocket and ligand embeddings for 72,199 protein-ligand complexes from the Protein Data Bank." Sole-author Data Descriptor under bioRxiv screening, 2026. Pipeline source: Zenodo, doi:[10.5281/zenodo.19954940](#).

## Skills

---

**Languages:** Python, Java, Kotlin, TypeScript, JavaScript, SQL

**Backend:** Node.js, Spring Boot, FastAPI, Express.js, REST, GraphQL, Kafka, Redis, SQS, Lambda

**Frontend / Mobile:** React, Next.js, Tailwind CSS, Jest, React Testing Library, Flutter, Firebase

**Infra / DevOps:** AWS (EC2, Lambda, S3, DynamoDB), Docker, GitHub Actions, PostgreSQL, MongoDB, Elasticsearch, observability

**ML / AI:** PyTorch, Hugging Face Transformers, LoRA / QLoRA, FAISS / HNSW, RAG, embeddings, RDKit, ESM-2, HumanEval, 4-bit / 8-bit quantization

## Honors

---

**Hack The Box, Top 100 Hall of Fame** ([verified badge](#)). Earned December 2021.