

# Himadri Mishra

Senior AI Engineer

Remote, India | +91-9918889939 | himadri.jobhunt@gmail.com  
linkedin.com/in/hmishra2250 | github.com/hmishra2250 | himadri.dev

## SUMMARY

---

Senior AI Engineer (IIT-BHU, 8 yrs) building production agentic AI platforms, verified LLM workflows, and ML infrastructure. Architected Knit 3.0's end-to-end research reporting platform: survey data to judged insights, charts, and PPTX decks, reducing 48–72 hour analyst turnaround to under 1 hour.

## EXPERIENCE

---

### Senior AI Engineer

*Remote, India*

*Knit*

May 2025 – April 2026

- Architected **Knit 3.0's end-to-end research reporting platform**: survey ingestion through verified insights, charts, memos, and PPTX decks, **reducing post-fielding turnaround from 48–72 hours to under 1 hour**.
- Built a **sandboxed insight engine** where LLMs generate and execute Python analyses across 30–50 tasks per report, with independent judge agents verifying each output against researcher-grade expectations.
- Designed **DAG-based orchestration** with dependency-aware parallel execution and a visualization pipeline producing 15–25 Highcharts charts per report with multi-threshold quality scoring.
- Unified reporting, questionnaire, ingestion, and deck agents on a **shared Python platform** with multi-provider LLM routing (GPT-4/5, Claude 4.x, Gemini 2.5/3.1 Pro), pgvector RAG, OpenTelemetry, Langfuse, SSE, and generated REST APIs.
- Owned technical direction as **principal architect and senior IC** for the India AI team, driving system-level architecture decisions across all AI-native market research workflows.

### Independent ML Engineer (Freelance)

*Remote, India*

*Stealth Product Team*

March 2025 – April 2025

- Built an **end-to-end text-to-vector-icon generation pipeline**, covering dataset creation, model benchmarking, LoRA fine-tuning, segmentation, and vectorization.
- Curated **18k vector icons** and generated high-quality descriptions through batch Gemini API calls; benchmarked proprietary and open-source text-to-image models including Flux and SDXL.
- Trained and evaluated thin-structure segmentation models including U-Net, U-Net++, U2-Net, and SegFormer, then built a graph-based skeleton-to-vector conversion workflow.

### Career Break

May 2024 – February 2025

*Personal sabbatical*

### Senior Research Engineer

*Remote, India*

*Epic! for Kids*

February 2023 – May 2024

- **Owned full ML pipeline and infrastructure** post-layoffs across discovery, recommendation, and search; led Elasticsearch autocomplete revamp, **outperforming prior solution in 80%+ cases**.
- Cut Docker build time by **50%**, reduced Kubernetes pod usage by **100x**, slashed spot instance errors by **99%**, cutting platform cost by **10x**.
- Launched ML features (book picker, performance-based carousels) via A/B testing, improving engagement; defined and executed ML roadmap resolving high-impact infrastructure issues.
- Drove cross-stack contributions across backend (PHP), frontend logging (Angular), and analytics including churn prediction.

## Senior Research Engineer

Tangible Play (Osmo)

Remote, India

September 2019 – February 2023

- Led **Computer Vision** across India and US teams: set technical direction, mentored engineers, and owned the CV roadmap for the Worksheets product.
- Raised CV accuracy in Worksheets from **93% to 98%** impacting millions of learners; built a real-time U-Net shaded-region detection model achieving **80% IoU**, deployed in Java.
- Led tracing the dots games for Pre-KG to Grade 3, **boosting engagement by 20%**; prototyped Byju's Board real-time teacher feedback system with **10fps detection**.
- Automated tagging workflows in Go and Dart, **cutting manual effort by 99%**.

## Deep Learning Engineer

Whodat (AR startup, acquired by Byju's)

Bangalore, India

July 2018 – August 2019

- Built fast ORB detector in C++, **20% faster than ORB-SLAM**; researched monocular depth estimation (DeMoN, MvDepthNet). *Team transitioned to Tangible Play/Osmo post-acquisition.*

## SKILLS

---

**Agentic AI and LLMs:** LLM Agents, Agentic Workflows, Multi-Agent Systems, RAG, LLM Evaluation, AI Guardrails, Prompt Engineering, LangChain, LlamaIndex, Hugging Face, BERT, OpenAI API (GPT-4/5), Anthropic API (Claude 4.x), Google Gemini 2.5/3.1 Pro

**ML and Data:** PyTorch, TensorFlow, Computer Vision, NLP, Survey Analytics, Market Research, MLOps, Statistics, Sentence Transformers

**Platform and Infrastructure:** FastAPI, REST APIs, SSE, PostgreSQL (pgvector), Redis, S3, Elasticsearch, PySpark, Highcharts, OpenTelemetry, Langfuse, E2B, Docker, Kubernetes, Google Cloud, CI/CD

**Languages:** Python, TypeScript, C++, Java, Go, Dart, PHP

**Web:** Django, Angular, Next.js, React

## PROJECTS AND AWARDS

---

**Kaggle: FIDE and Google Efficient Chess AI Challenge (2025):** Secured **top 6% globally**, Rank 60 out of 1120.

**Research Intern, UC Berkeley (2017):** Neural programmer-interpreters under Professor Dawn Song. *SN Bose Scholar Program Awardee.*

**Software Engineering Intern, Microsoft (2016):** Dialog-based chatbots for the Incentive Compensation team.

**Earlier:** Top 8% Kaggle Allstate Claims Severity (2017); ACM ICPC Amritapuri Regional (2017); HP Think-A-Thon (2016).

## EDUCATION

---

**Indian Institute of Technology (IIT-BHU), Varanasi**  
Integrated Dual Degree in Computer Science and Engineering  
GPA: **9.28/10**

2013 – 2018