INBASEKARAN PERUMAL

+91 8660467304 | inba2002.p@gmail.com
linkedin.com/in/inbasPeru | github.com/the-pinbo | inbasperu.github.io

PROFESSIONAL SUMMARY

Software engineer taking AI features from ambiguous product needs to production, with ownership across solution design, backend implementation, evaluation, and deployment. Combines product judgment with hands-on engineering to build useful, reliable AI capabilities.

TECHNICAL SKILLS

Languages: C++, Python, TypeScript
AI and ML: Agent harness engineering, RAG, LLM evaluation, MCP, Arize Phoenix, LangChain, Deep Agents
Backend and Infrastructure: Node.js, FastAPI, Elasticsearch, Postgres, Kubernetes

WORK EXPERIENCE

RingCentral

Software Development Engineer II (Jul 2026–Present)

- Helped reduce monthly AI spend 52% ($16.2K $\rightarrow$ $7.8K) while supporting higher delivery volume through shared planning context, role-specific agent skills and workflows, persistent memory, and team training.
- Cut estimated diagnostic time from 30 to 5 minutes and helped the team resolve 123 Jira bugs in three weeks through agentic debugging workflows and an internal MCP unifying observability context.
- Enabled self-service reporting across 23 conversation-intelligence entities, bypassing month-long engineering cycles; designed the MCP query language and evaluated reporting accuracy through multi-turn tests.

Software Development Engineer I (Jul 2024–Jun 2026)

- Proposed and shipped natural language filtering using RAG to discover and apply 40+ predefined filters, with filter activity up 13% in the first month. Used production traces and evaluations to route about 40% of queries through deterministic logic, cutting their latency from 5 seconds to under 1 second.
- Reduced average call listing latency from over 5 seconds to approximately 1.2 seconds by redesigning Elasticsearch queries and mappings and improving caching. Modernized application APIs and contributed to quarterly index partitioning and sharding to support ACE's customer base growing 60% YoY.
- Independently designed and built the Insights analytics backend across 40+ signal types in a service generating 4M+ signals/day. Separated mutable classifications into a dedicated Elasticsearch index with server-side aggregation, enabling daily category updates without rewriting historical signals.
- Enabled cross-language search across 5 additional European languages through translated query and synonym expansion, achieving 85% accuracy on 80 synthetic benchmark queries. Added on-demand translation middleware for dynamic content, avoiding storage of separate translations for each language.

The Linux Foundation Mentorship Program

Open Source Developer, SoftPosit for RISC-V (Mar 2025–May 2025)

- Enabled broader evaluation of posit arithmetic, an alternative to floating point, for RISC-V by implementing missing 8- and 16-bit math functions in C++ and adding unit tests for each.

Google Summer of Code

Open Source Developer, Intel OpenVINO Toolkit (May 2024–Aug 2024)

- Built VisionGuard, a gaze-tracking app showcasing on-device AI on Intel NPUs, with 46% lower measured power consumption than GPU execution (14 W vs. 26 W); featured in Awesome OpenVINO.

Indian Institute of Science

Research Intern, Spectrum Lab (Apr 2024–Jul 2024)

- Improved signal recovery from noisy measurements by implementing and benchmarking amplitude debiasing in LISTA, raising mean reconstruction SNR by 6.2 dB (26.4 $\rightarrow$ 32.6 dB).

WORK EXPERIENCE CONTINUED

Goldman Sachs

Summer Analyst Intern, Cards Marcus Data Engineering (May 2023–Jul 2023)

- Saved team leads 30 minutes daily by automating progress and service-target reporting for 15 workflows using PySpark on Databricks.

Determinant Studios

Software Engineering Intern (Dec 2022–Feb 2023)

- Enabled digital inventory tracking for shopkeepers with limited literacy by prototyping spaCy entity extraction for voice input, reducing entry time about 35% in tests; later incorporated into Vidhi, now supporting regional languages with 1K+ Google Play downloads.

PUBLICATIONS AND PATENTS

Some Intriguing Observations on the Learnt Matrices in Deep Unfolded Networks
Co-author | IEEE ICASSP 2025 Proceedings

Evaluating Numerical Formats for HPC Using Fractal Benchmarks
Co-author | CoNGA 2025 Proceedings, forthcoming

Voice Assisted System, Device and Method for Inventory Management
Named inventor | Indian Patent No. 576831 | Granted 2025

AWARDS

Best Debut Award, RingCentral
Exemplary performance in Q4 2024; awarded May 2025

EDUCATION

National Institute of Technology Karnataka, Surathkal

B.Tech, Electronics and Communication Engineering | CGPA: 9.19/10 | Class rank: 4/138 | 2020–2024