

Ayon Sarkar

+91-81002-11809 [in](#) [G+](#) [M](#) [T](#) ayon380@icloud.com

EXPERIENCE

- Nucleus AI** July 2025 – April 2026
Software Development Engineer — AI Infrastructure San Francisco, USA (Remote)
 - Built and operated the **training data pipeline** for a production AI system, migrating **2B+ images** across Azure tenants using **Azure Data Factory**, resolving cross-tenant throughput bottlenecks and cutting storage costs.
 - Architected a **distributed data harvesting system** on **AWS Lambda + SQS** to collect **TBs of multimodal training data**; implemented adaptive backoff and rate-limit evasion strategies to maximize ingestion throughput.
- Nucleus AI** August 2024 – July 2025
SDE Intern — AI Product Frontend San Francisco, USA (Remote)
 - Engineered the **frontend for a consumer-facing AI chat product** using **Next.js, Radix UI, and ShadCN**, achieving a **99/100** Google Lighthouse score under production load.
 - Reduced **idle memory footprint to 30MB** (peak 100MB during active AI inference sessions) through aggressive **dynamic component loading** and optimized streaming UI rendering — critical for low-end device support.
- Agumentik Group** June 2023 – August 2023
Full Stack Developer Intern Remote
 - Delivered **10+ core features** (search, cart, authentication) for the Markito e-commerce platform, optimizing to a **90+ Lighthouse score** and measurably improving user session times.

SKILLS

Languages & Frameworks:	NextJs, React, TypeScript, Node.js , Python , C++, Go, SQL, Shell (Bash)
Backend & Databases:	Express, MongoDB, Redis, PostgreSQL, REST APIs
Infrastructure & Cloud:	AWS (Lambda, SQS, SNS, CloudWatch, API Gateway), Docker, Kubernetes (EKS), Kafka, Terraform
Concepts & Tools:	System Design, Distributed Systems, Event-Driven Architecture, SOLID, Microservices

PROJECTS

- Aurex: High-Performance C++ Order Matching Engine** Sep 2025 – Present
Tools: C++, Redis, PostgreSQL, Next.js, WebSocket [GitHub]
 - Engineered a **lock-free C++ matching core** capable of **50K+ orders/sec** at **p50 latency of 2μs** using custom multi-threading, lock-free queues, and **CPU affinity optimizations** — directly applicable to low-latency AI inference serving.
 - Designed a custom **raw TCP binary protocol** with **Nagle’s algorithm disabled** and optimized socket buffer tuning, achieving **~500μs** round-trip latency vs. **gRPC’s 1–2ms** — a **4x improvement** on the critical matching path.
 - Streamed real-time market data to a **Next.js dashboard** via **gRPC**, maintaining high-throughput data conditions.
- Aptinova — AI-Powered Recruitment Platform** Jan 2025 – May 2025
Tools: Next.js, Gemini AI, PostgreSQL, FastAPI [Live Link]
 - Integrated **Google Gemini** to build an **AI resume parser and candidate-matching pipeline**, automating structured data extraction from unstructured documents.
 - Built a **Secure Assessment Module** with **8+ real-time proctoring signals** (focus-loss detection, tab-switch monitoring) using browser-native APIs and WebSockets.
- Iris — macOS Screen Health Manager** Jan 2026 – Present
Tools: Swift, SwiftUI, AppKit [GitHub]
 - Building a **native macOS menu-bar app** to combat computer vision syndrome; engineered **system-wide interrupt overlays** using AppKit with non-bypassable timer enforcement.

RESEARCH & PROFESSIONAL SERVICE

- Enhanced Environmental Perception for Visually Impaired** [Paper Link]
Published in IEEE Xplore
 - Developed a **real-time object detection system** using **YOLOv4**, achieving **83% accuracy** with sub-100ms auditory feedback latency for assistive use.
- Conference Reviewer: IEEE IATMSI-2026** [Certificate]
4th IEEE International Conference on Technology and Management
 - Appointed as **technical reviewer** for research papers on technology and management innovation.

EDUCATION

- SRM University AP** August 2021 – May 2025
B.Tech in Computer Science — Cloud Computing Specialization Amaravati, India
 - **CGPA: 8.2/10**