KHAGENDRA KHATRI

(862)-276-5565 | Montclair, NJ | realkhagendra@gmail.com | Website | LinkedIn

EDUCATION

Caldwell University Caldwell, NJ
Bachelor of Science in Computer Science | GPA: 3.82/4.00 Expected May 2025

EXPERIENCE

Software Developer Intern Caldwell, NJ
Digital Coach Oct 2024 - Present

- Architected and deployed a distributed e-learning platform using Next.js and React with server-side rendering, achieving 20% reduced latency and supporting 500+ concurrent users with real-time sentiment analysis.
- Designed a microservices architecture using Firebase Cloud Functions and Pub/Sub, implementing circuit breakers and retry mechanisms to achieve 99.9% system availability
- Engineered a real-time feedback system using WebSocket connections and Redis pub/sub, processing 100+ simultaneous mock interviews with <100ms latency.

AI Engineer Intern Remote
Radical AI Jun 2024 - Aug 2024

- Designed and implemented a distributed LLM orchestration system handling 100+ requests/minute using OpenAI and Gemini APIs, with automatic failover and load balancing achieving 99.95% uptime.
- Built a fault-tolerant data pipeline using Apache Airflow and Docker, processing 10GB+ of training data with automated validation and monitoring, reducing data preprocessing time by 60%.
- Implemented semantic caching and query optimization strategies, reducing API costs by 45% and average response time by 35%.

Machine Learning Researcher Caldwell, NJ
Caldwell University May 2024 - Aug 2024

- Invented and evaluated Recurrent Symbolic Networks architecture to tackle the vanishing gradient problem in RNNs, enhancing performance by 17% and interpretability for sequential data analysis.
- Collaborated with professors and fellow researchers to refine research methodologies, analyze results, and publish findings in the **28th** annual conference of the Consortium for Computing Sciences.

PROJECTS

Crystal Math | Flask, React.js, Manim, PostgreSQL, AWS, VerbWire, Terraform | DoraHackPost

- Architected and deployed a scalable mathematical visualization platform serving 100+ monthly users, implementing microservices architecture orchestrated via AWS ECS for automated scaling and 99.9% uptime.
- Designed an event-driven architecture using AWS Lambda and SQS for asynchronous processing of math animations, reducing rendering time by 60% and handling 100+ concurrent user requests.
- Optimized cloud costs by 40% through infrastructure-as-code using Terraform, implementing auto-scaling policies and spot instances for compute-intensive rendering tasks.

E-Graveyard | Django, React.js, PostgreSQL, Redis, Celery, LangChain, Elevenlabs | DevPost

- Designed and implemented a high-throughput conversational AI platform handling 100+ concurrent users with <100ms response time, using Django Channels for WebSocket connections.
- Engineered a distributed task processing system using Celery and Redis Queue, achieving 5x improvement in voice generation throughput and reducing API costs by 50% through intelligent caching and batch processing.
- Implemented a fault-tolerant vector search system using PostgreSQL pg_vector and HNSW indexing, achieving sub-second query performance across 20k+ embeddings with 98% accuracy.

SKILLS

- **Languages:** C#, Python, JavaScript, TypeScript, SQL, Java, C, Visual Basic, C++, OCaml
- **Frameworks**: Django, Flask, ASP.NET MVC Controller, .NET Framework, React, Next
- **Libraries:** Tensorflow, Pytorch, Scikit-learn, Matplotlib, Pandas, Numpy, NLT

AWARDS & ACHIEVEMENTS

1st Place, Best Use of Terraform, and Best Use of Verbwire at HackPrinceton Fall 2024 (500+ participants); 2nd Place at HopHacks Fall 2024 (100+ teams); Best Startup to Launch at HackRU Fall 2024 (400+ hackers); and Best Use of Midnight at HackNJIT Fall 2024 (500+ participants); Awarded Presidential Scholarship; Dean's List honoree (2021–2024); Top 5 Finalist, National Olympiad in Informatics (2020).