HARSHIT JAIN

+1 (582) 203-9755 ◇ harshitj.cs@gmail.com ◇ linkedin.com/in/harshitjain17 ◇ github.com/harshitjain17

EDUCATION

The Pennsylvania State University, University Park PA Expected Graduation: Dec 2024
Bachelor of Science in Computer Science
GPA: 3.7/4.0 | Teaching Assistant @CodePath | AlgoPSU Captain @ACM | Recommendations | Projects | Resident Assistant
Coursework: Data Structures and Algorithms, Operating Systems, Systems Programming, CodePath: (Intermediate+Advanced) Software Engineering, Object-Oriented Programming, Linear Programming, Theory of Computation, Database Systems
Certifications: AWS, Machine Learning, Advanced Learning Algorithms (Deep Learning), Generative AI with LLMs

TECHNICAL SKILLS

Programming Languages: Python, Java, C/C++, JavaScript/TypeScript, SQL, Shell scripting, MATLAB
Frameworks: TensorFlow, Scikit-Learn, LangChain, Streamlit, Node.js, React.js, Next.js, GraphQL, RESTful APIs, Git
Softwares: AWS, GCP, Azure, MS SQL Server/PostgreSQL, Linux/UNIX, Docker, SonarQube, Postman, Swagger, JIRA

WORK EXPERIENCE

Software Engineer Intern Aug 2024 - Present
Materials Research Institute (2DCC-MIP Team) University Park, PA

- Co-built a scalable platform (Next.js) to enable low-latency ML inference using AWS to handle real-time video processing
- Optimized video retrieval by designing GraphQL APIs to enhance data handling via DynamoDB integration for CRUD Ops
- Enabled secure CRUD Ops in a Next.js app by connecting AWS API Gateway to Lambda, authenticated with AWS Cognito
- Deployed an agent workflow (Python) with LangChain and utilized Pinecone to convert video metadata to vector embeddings
- Technologies Used: Python, Amazon Web Services (AWS), LangChain, Pinecone, Next.js, GraphQL, REST APIs

Software Engineer Intern May 2024 - Aug 2024
Hughes Network Systems, LLC (Aeronautical Team) Washington, DC

- Automated AWS to BigQuery pipeline (Python) with strict row-level validation and reduced manual intervention by 95%
- Developed Python scripts and optimized SQL views to improve flight performance metrics for 61k+ flights' data per month
- Designed an automated billing script on GCP for dynamic invoicing of flight performance and reduced manual overhead
- Technologies Used: Python, Google Cloud Platform (GCP: Cloud Functions, BigQuery, Pub/Sub), SQL, Agile/Scrum

Software Engineer Intern Jan 2024 - May 2024
Materials Research Institute (2DCC-MIP Team) University Park, PA

- Integrated LLMs (GPT-4o, Claude 3) into a LangChain-based chatbot framework for answer retrieval via a RAG model
- Built a serverless Next.js app integrated with AWS and streamlined RESTful APIs for high scalability and CI/CD workflows
- Developed a fully automated AWS Lambda pipeline for video processing, including transcription and AI-generated metadata
- Configured AWS Rekognition to enhance speaker identification, improve transcription, and automate chapter generation
- Technologies Used: Python, Amazon Web Services (AWS), LangChain, Next.js, RESTful APIs, LLMs

Software Engineer Co-op May 2023 - Dec 2023
VIAVI Solutions Inc. (PNT Team) Washington, DC

- Built distributed test suites in Python and C++ for PNT products, enabling parallel testing and boosting reliability by 30%
- Boosted system performance by optimizing multi-tiered architecture with microservices and debugging SCPI protocol issues
- Executed 35+ SCPI-driven long-term tests on core devices to ensure compliance with release-level quality standards
- Technologies Used: Python, C/C++, Linux, SCPI Protocol, Bitbucket, Confluence, SonarQube, Git, JIRA, Agile/Scrum

Software Engineer Intern May 2022 - May 2023
Materials Research Institute (2DCC-MIP Team) University Park, PA

- Engineered a scalable React.js app with 50+ user-facing features across 20+ components, and employed Jest for unit testing
- Integrated MS SQL Server to manage data from 500+ instruments, and organized it across 18+ BCNF-normalized tables
- Built a Python-based ETL pipeline to automate data extraction and integrated the data with a cloud-based data warehouse
- Research: Re-architected a Python library for the Raman Fitting Model (try here) to implement precise spectra deconvolution algorithms with automated data preprocessing, fitting, and export functions, reducing analysis time by 40% (validated)
- Technologies Used: Python, JavaScript, React.js, Node.js, MS SQL Server, RESTful APIs, Git, HTML/CSS, JIRA

PROJECTS

Dynamic Memory Allocator [C/C++] Jan 2024 - Feb 2024

- Designed custom malloc, free, realloc; segregated free lists and footer optimization to improve memory management
- Achieved a utilization score of 69% and benchmark throughput at 100% across diverse computing environments