VANSH BATAVIYA

US Citizen | bataviyavansh@gmail.com | in/vansh-bataviya | github.com/vanshb03 | vansh.nyc

Education

Hunter College, City University of New York May 2027
Bachelor of Arts in Computer Science *Manhattan, NY*
- **Relevant Coursework:** Data Structures, Algorithms, Object Oriented Programming, Software Engineering, Computer Architecture, Software Analysis and Design, Discrete Mathematics, Linear Algebra, Statistics

Experience

Capital One May 2025 – August 2025
Software Engineering Intern *Manhattan, NY*
- Built Card Statements Wrapped for **140M** customers, processing **10.8B** annual transactions through AWS and Python pipeline.
- Designed automated ETL system handling **375B** payment decisions annually, implementing PySpark workflows with DynamoDB.
- Developed daily batch processing pipeline transforming raw transaction data, enabling real-time API queries for mobile integration.

Cvrve August 2024 – April 2025
Co-Founder *Manhattan, NY*
- Engineered ClickHouse architecture processing **5.2M+** job interactions, reducing storage by **50%** through columnar compression.
- Developed open-source repositories reaching **7,000+** GitHub stars with automated data collection for nationwide role discovery.
- Built multi-threaded ETL processes handling **10K+** concurrent API requests per minute with **99.84%** uptime for real-time aggregation.

Capital One May 2024 – August 2024
Software Engineering Intern *Mclean, VA*
- Accelerated data processing by **46%** via Python SDK integrating PostgreSQL databases with OneStream API using AWS Lambda.
- Standardized data ingestion for Accounts and Matrix datasets from PostgreSQL and AWS S3 with batch-optimized streaming APIs.
- Delivered production code with **92%** test coverage implementing automated unit testing and Jenkins CI/CD pipelines for deployment.

Terascale All-Sensing Research Studio (TARS) November 2022 – May 2023
Undergraduate Researcher *Potsdam, NY*
- Contributed to DeepJoin, an automated approach using neural networks to generate high-resolution repairs for fractured objects.
- Embedded Python for shape representations, improving deconstruction accuracy of fractured objects by **35%** over baselines.
- Crafted restoration structures for real-world objects with **84%** accuracy using synthetic fracture training, enabling 3D-printed repairs.

Projects

YaleEd Yale Hackathon ’24
- Implemented lightweight BERT text analysis to extract keywords from task descriptions, creating an automated priority system.
- Implemented secure OAuth authentication using Google accounts with NextAuth and Google Identity Platform.
- Orchestrated scalable data ingestion and transformation pipelines, enhancing the dashboard's real-time analytics for 500+ users.

Henlingo HenHacks ’24
- Developed award-winning web application that utilized image recognition to teach American Sign Language (ASL).
- Configured AWS Lambda with API Gateway to efficiently process recognition requests, enabling real-time feedback for ASL learners.
- Engineered responsive layouts with TypeScript and TensorFlow to interpret user gestures and provide immediate feedback.

Mindguard Columbia Univ. Hackathon ’24
- Engineered efficient machine learning algorithms using BiLSTM and BERT to filter triggering content from YouTube feeds.
- Leveraged AWS Sagemaker to train customized NLP models for contextual understanding of potentially harmful video content.
- Implemented automated testing frameworks in AWS CodeBuild to validate model performance before presentation demo.

academ.ai HackRPI X
- Designed an educational tool leveraging GPT-4 to generate personalized study schedules, enhancing individual engagement.
- Developed validation tests using Google Data Validation Tool (DVT) for Cloud SQL reliability and functionality.
- Utilized PyQt6 to create an intuitive graphical interface (GUI) for academ.ai, enhancing user accessibility.

Skills

Languages Python, JavaScript, TypeScript, Golang, Swift, Java, C, C++, SQL, HTML/CSS, R
Frameworks Next.js, React, Node.js, FastAPI, Tailwind, Angular, Bootstrap, Express.js, Flask
Tools Google Cloud, AWS EC2, AWS Lambda, AWS RDS, Jenkins, Postgres, Kafka, Docker, Terraform, MongoDB, GraphQL, Git
Python Libraries NumPy, Beautiful Soup, Matplotlib, OpenCV, Pandas, Scikit-learn, PyQt6, Tensorflow, Playwright
Concepts Backend, Distributed Systems, Microservices, Networking, Linux, Software Engineering, System Design, REST APIs