

Aryan Singh

(848) 327-5741 | aryansingh080910@gmail.com | [linkedin.com/in/aryansingh123](https://www.linkedin.com/in/aryansingh123) | github.com/AryanSingh103

EDUCATION

Rutgers University

New Brunswick, NJ

B.S. in Computer Science; B.A. in Cognitive Science (GPA: 3.73 | Dean's List)

Expected May 2028

- **Relevant Coursework:** Data Structures, Linear Algebra & Optimization, Computer Architecture, Discrete Structures, Principles of Programming Languages, Multivariable Calculus, Intro to AI, Distributed Systems

EXPERIENCE

AI Engineering Intern

Nov 2026

BalanX Bio

Remote

- Role focused on designing, testing, and refining AI personalization models, and collaborating with data and engineering teams to deploy AI features into production.

Lead Software Engineer

Jan 2026 – Present

HCI Lab for Health & Sustainability, Rutgers University

New Brunswick, NJ

- Engineered and deployed What's Happening Around Me, a full-stack Next.js/TypeScript air-quality platform using weather and OpenAI APIs, leading a 25+ participant study whose findings were co-authored into a paper submitted at ACM CHI 2027, the premier venue in human-computer interaction research.
- Identified a gap in ActivityWatch's browser-tracking capabilities, then merged a custom activity classifier (Python) via a PR review into the 19k-star open-source project, enabling browsing-activity data for a lab research study.
- Served as the lab's primary software engineer, applying communication skills to translate requirements from 15+ researchers into shipped features including interactive mapping, location comparison, and a streaming AI chatbot.

Machine Learning Extern

June 2026 – Aug 2026

Collaborative Solutions, LLC

Piscataway, NJ

- Led team's migration from YOLOv5 to YOLOv11 for a model detecting visual distractions at railroad grade crossings, building training infrastructure across GitHub, Drive, and Colab for 15+ model training experiments
- Fine-tuned YOLOv11 models via transfer learning from pretrained COCO weights, testing configurations to increase model accuracy by 33% over the prior team's benchmark on data (mAP50: 0.406 → 0.606).
- Diagnosed a bug silently shrinking training images before the model saw them, then fixed the pipeline to preserve image quality, lifting model accuracy 14% and thin-object accuracy 111% (mAP50 +14%, thin-object AP +111%).

PROJECTS

AccessAdvisor – WCAG Auditor | *Python, Next.js, TypeScript, FastAPI, Claude API, LlamaIndex, ChromaDB*

- Built AccessAdvisor, a full-stack AI tool that scans GitHub pull requests, live URLs, and entire repositories for WCAG 2.1 violations, grounding findings in 650 chunked spec passages to eliminate hallucinations for upto 99%.
- Built a filtering system that skips irrelevant files before running any AI analysis, cutting out 73.5% of files across 5 real codebases and reducing analysis cost by up to 98.5% on a real 30-file code review.
- Built a 30-case evaluation harness against all 78 WCAG criteria and found retrieval made results worse, catching violations 90% of the time without it versus 60% with it, traced to a bug dropping the right answer in 1 of 3 cases.

AuctionEngine – Distributed Bidding System | *Go, PostgreSQL, Kafka, Redis, Docker, WebSockets, Prometheus*

- Built a real-time auction engine in Go and PostgreSQL sustaining throughput of 1,100 bids per second on a single live auction at 3.3 ms p99 latency, holding correctness across 1,000 concurrent bidders with zero lost updates.
- Guaranteed each auction winner is charged exactly once despite duplicate Kafka messages and mid-charge crashes, using a transactional outbox and idempotency keys, with zero double charges at 50% simulated payment failure.
- Provisioned AWS ECS, RDS and ALB entirely as code with Terraform, deploying every commit through GitHub Actions with health-checked rolling updates and no downtime.

NeuroDriveHelper – AI Reading Assistant | *JavaScript, FastAPI, Hugging Face Transformers, ElevenLabs API*

- Built and shipped a Chrome extension (Manifest V3) that captures on-page text and simplifies it in real time via a Dockerized FastAPI backend running Hugging Face Transformer models.
- Integrated ElevenLabs text-to-speech, a dyslexia-friendly rendering mode, and a distraction-reducing focus mode, giving neurodiverse users three complementary ways to process on-page content.

SKILLS

Languages: Python, Go, C/C++, TypeScript, JavaScript, Java, SQL, Scheme, OCaml

Frameworks: FastAPI, Flask, Next.js, Node.js, React Native, Tailwind CSS, Spring Boot

Developer Tools: Docker, AWS, Terraform, PostgreSQL, Redis, Kafka, Prometheus, Grafana, GitHub Actions

AI/ML: PyTorch, scikit-learn, TensorFlow.js, LangChain, LlamaIndex, ChromaDB, OpenAI/Claude APIs, CNNs,

OCR, Transformers, Reinforcement Learning, Transfer Learning, Deep Learning, Pandas, NumPy