

# Haradeep Puneti

804-767-9470 | haradeep.puneti@gmail.com | linkedin.com/in/haradeep1 | github.com/hpuneti1

## EDUCATION

### Virginia Tech

*Bachelor of Engineering in Computer Science - GPA: 3.92*

**Relevant Coursework:** Data Structures & Algorithms, Linear Algebra, Computer Organization, Applied Combinatorics, Software Design, Machine Learning

### Virginia Tech

*Master of Engineering - Data Analytics and Artificial Intelligence*

Blacksburg, VA

Aug. 2024 – Dec. 2026

Blacksburg, VA

Dec. 2027

## TECHNICAL SKILLS

**Languages:** Java, Python, SQL (Postgres), JavaScript, HTML, CSS, C

**Frameworks/Libraries:** React, Node.js, Express.js, Flask, OpenCV, YOLO, Tesseract OCR

**Databases/Tools:** PostgreSQL, Weaviate, Neo4j, ChromaDB, Git, Docker, Power BI, Streamlit

## EXPERIENCE

### Software Engineer Intern

May 2026 - Present

*CACI*

*Ashburn, VA*

- Building a PyTorch computer vision pipeline pairing YOLO detection with a ResNet-50 re-identification model to classify vehicle make, model, and color and track vehicles across multiple overlapping border camera feeds
- Enhancing CI/CD pipelines to automate testing and deployment, while prototyping a multi-agent AI system to further streamline build and review workflows

### Undergraduate Researcher

September 2025 - Present

*College of Engineering, Virginia Tech*

*Blacksburg, VA*

- Built an end-to-end **NLP benchmarking pipeline** comparing LLM-based entity extraction against traditional and neural baselines (**spaCy**, **GLiNER**, **KeyBERT**) across various scientific text domains
- Designed a **reproducible evaluation framework** with fuzzy matching, precision/recall scoring, and multi-replication statistical analysis to assess zero-shot LLM performance on qualitative text
- Authored findings for conference publication examining LLM capability boundaries in domain-specific concept identification, with a custom full-stack annotation and analysis webapp

### Software Engineering Intern

May 2025 – December 2025

*AIRCO*

*Brooklyn, NY - In Person/Hybrid*

- Developed automated BFD and PFD translation/generation system using **Python** and **Aspen APIs**, reducing manual diagram translation time by **15%** for renewable fuel process design workflows
- Built computer vision pipeline using **OpenCV**, **YOLO**, and **Tesseract** to detect and classify components in PID diagrams across **50+ diagrams**, outputting structured JSON for downstream processing
- Designed an AI agent system automating engineering dataset management within Aspen Basic Engineering, cutting manual data entry time by **25%**

### Data Science Intern

June 2023 – August 2023

*ITVISION360*

*Plano, TX - Virtual*

- Authored **15+ stored procedures and views** in **PostgreSQL** to support BI reporting, processing datasets of 1M+ rows and reducing query latency by **30%**
- Built **Power BI** data pipelines automating previously manual reporting workflows, delivering **20+ dashboards** to stakeholders and reducing weekly report prep time by **4 hours**

## PROJECTS

**Satori** | *React, Node.js, PostgreSQL, Weaviate, Auth0*

- Built a full-stack SAT prep platform serving **30+ users** across old SAT, PSAT, and Digital SAT formats, featuring AI-powered semantic search over **2000+ practice questions** via OpenAI embeddings and JWT-authenticated progress and subscription tracking

**Green Market Farms** | *Next.js, TypeScript, Supabase, PostgreSQL, Stripe, OpenAI, Gemini - VT GDG Hackathon*

- Delivered a production farm e-commerce platform serving **500+ users** for a real client with **Stripe** payments, **pgvector** semantic search, and an AI voice assistant for hands-free inventory and order management
- Engineered full-stack order lifecycle with webhook-driven confirmation, real-time SMS and email notifications, and an admin dashboard with sales analytics actively used by the organization

**Vocet** | *Next.js, TypeScript, AWS Bedrock, AWS S3, AWS Lambda, Clerk, Upstash, OpenAI*

- Building an AI voice interviewer for system design prep featuring real-time turn-taking with silence detection, a hybrid whiteboard (tldraw + vision API), and a curated question bank with AI-generated per-session variations
- Architecting a multi-model **AWS Bedrock** pipeline that routes between a fast model for intent classification and a strong model for evaluation; offloading post-session feedback report generation to **AWS Lambda** with session audio and whiteboard snapshots persisted to **AWS S3**

**IMC Prosperity 2025** | *Algorithmic trading competition by IMC Trading 204th US / 772nd Global of 18,803*