

# Soham Kundu

skundu6@wisc.edu • sohamkundu.dev • linkedin.com/in/sohamkundu27 • (763) 406-1195 • US Citizen

## EDUCATION

### University of Wisconsin-Madison

BS in Computer Engineering and Computer Science

GPA: 3.7/4.0

Expected May 2027

- MadData25 **Hackathon 1st Place** Winner (30+ Teams, 120+ Participants), Top 3 at University Madness Startup Pitch Competition (9 Universities, 25+ Teams), Eagan Foundation Scholarship, Dean's Honors List x2
- Coursework- Data Structures & Algorithms, Object Oriented Programming, Computer Architecture, Machine Learning, Big Data Systems, Linear Algebra and Differential Equations, Discrete Math, Multivariable Calc

## EXPERIENCE

### Blue Cross Blue Shield - AI/Software Engineering Intern

May 2025 – Present

- Built **agentic-AI** prototypes with Azure AI, integrating .NET/**Python** services for real-time prior-authorization support
- Pioneered real-time knowledge retrieval pipeline combining keyword & semantic search, cutting Customer Service Rep search time by 90% and accelerating patient access to critical healthcare info, **impacting 3M+ members**
- Implemented **LangChain** to orchestrate and deploy Large Language Model-based knowledge retrieval for live CSR use

### FiPet - Lead SWE/Co-Founder (Part-time) | [fipet.dev](#)

May 2025 – Present

- Architecting a mobile app platform with AI-powered, gamified financial education modules tailored for Gen Z audiences
- Built full-stack **Firebase** and **React Native** app integrating **ML-driven personalization engines** for adaptive learning
- Executed and designed full intern hiring pipeline, interviewing and onboarding 10+ engineering candidates
- Led 10-person development team across MVP roadmap and intern management; resulting in **200+ user** waitlist

### Machine Learning for NLOS Imaging Research - Under Prof. Andreas Velten

January 2025 – May 2025

- Engineered python software for embedded NLOS hardware integrating lasers and SPADs for real-time photon data capture
- Designed and trained **CNNs** in Python and MATLAB to reconstruct occluded objects from sparse time-of-flight datasets
- Optimized neural network inference, reducing computational overhead by 20% for deployment on edge devices

### Outlier AI - Software Engineer

May 2024 – August 2024

- Optimized **LLM** code for 200+ tasks across 3 projects, improving accuracy by 25% while preserving efficiency and edge cases
- Debugged and refactored codebases (**Java, Python, JS**), integrating **RL** to fine-tune multi-turn LLM reasoning
- Collaborated with research, engineering, and product teams to optimize model outputs and system accuracy

### Wisconsin Autonomous - Software Engineer

September 2024 – Present

- Built ML algorithms with **OpenCV** for real-time lane, cone, and boundary detection in autonomous vehicles
- Integrated perception module into software pipeline coordinating sensor fusion, control systems, and vehicle decision-making
- Collaborated with a 30+ person software team, managing version control, code reviews, and **CI/CD** workflows in **Git**

## PROJECTS

### ClaimReady | [claim-ready.vercel.app](#) | [Github](#)

- Co-founded a 2x award-winning AI startup web app to cut insurance claim valuation time from 20+ hrs to minutes
- Utilized YOLO11 for detection, Gemini API for identification/pricing, and Supabase for auth; processed 1200+ images
- Deployed via **Docker** on an **AWS EC2** instance; scaled to support our **350+ users**
- Recognized as “investor-ready” by judges; acknowledged by YC's Jared Friedman, Ankit Gupta, and UW-Madison CS Dept

### Hyperacing | [hyperacing.us](#) | [Github](#)

- Built and scaled a Go backend for a Formula 1 analytics platform with AI insights and live telemetry to **200+ users**
- Integrated Firebase Auth for secure login and session management, and used to F1 WebSockets to stream real-time race data
- Engineered a Next.js frontend with dynamic dashboards, a live F1 news feed, and live analytics

### Voice-Controlled MCP Robot Arm | [Demo](#) | [Github](#)

- Built a **full-stack** voice interface to control a robot arm by transcribing real-time speech into structured robot commands
- Engineered an **LLM** task planner to convert prompts like “clear the table” into JSON robot actions executed via MCP
- Validated the entire pipeline with async audio input and physics-based simulation, ensuring robust task completion

### Greenhouse IoT Monitoring System | [Github](#)

- Built IoT greenhouse system tracking temp, humidity, and brightness on Raspberry Pi in Python, optimizing GPIO control
- Configured Azure IoT Hub and **REST APIs** for real-time transmission of over 10K data points daily
- Deployed a **Dockerized Django-React** app on Azure App Services, enabling remote monitoring by parents in India

## TECHNICAL SKILLS

**Languages** - Go, Java, JavaScript, Python, SQL

**Frameworks/Libraries** - Django, Flask, Matplotlib, NumPy, OpenCV, Pandas, Pytorch, Sklearn, TensorFlow, .NET

**Developer Tools** - AWS, Azure, CI/CD, Docker, Git/GitHub, Google Cloud, Linux, REST APIs, Test-Driven Development