Sid Su

sidsu.com
sidsu7@gmail.com
linkedin.com/in/sid-su

Machine Learning Engineer, Full Stack Developer and DevOps Engineer with a strong foundation in *nix systems, coding and mathematics. Specializing in AI, automated pipelines and distributed systems, with a passion for leveraging emerging technologies to deliver reliable solutions

IBM | Office of the CTO | Machine Learning Engineer | May 2023–PRESENT

(Classified) **DoD CDAO "Project Harbinger" (Jan. 2025–Present):** A machine learning image segmentation project. Worked as a Machine Learning Engineer focusing on the preprocessing and postprocessing of data
Tools Used: Python, PyTorch, Scikit-Learn, Pandas, Numpy, MySQL, Bash, Makefile, Docker

(Herndon, VA) **DoD DIU Digital On-Ramp (Oct. 2024–Dec. 2024): (Contract Won)** A Generative AI chatbot that helps the DoD find vendors and products, and vendors find contracts for their products. Worked as a Full Stack developer on the GenAI team to create the chatbot and integrate it into the Appian platform
- **Chatbot** – Developed the chatbot response by aggregating and synthesizing government and private data
- **Tool Provider Integration** – Integrated several 3rd party tool provider APIs with RAG (retrieval-augmented generation) to generate better responses
- **Source Reporting** – Developed a pipeline to display the original data sources to the user
- **Persona UI** – Enabled a tailored, personalized "persona" user experience by dynamically adapting the UI, prefabricated prompts and chat responses based on user profiles and company attributes
- **Optimization** – Sped up chats by 30% with concurrency tuning. Sped up API responses by 50% with numerical interpolation and prompt engineering
Tools Used: JavaScript, Python, React.js, MUI, Node.js, Express.js, FastAPI, AWS Lambda, IBM Golden Retriever, CrewAI, LangChain, AWS Bedrock, MongoDB, AWS DocumentDB, PostgreSQL, Microsoft Excel

(Reston, VA) **The EchoNet Prod Team (Jun. 2024–Oct. 2024):** A Deep Learning AI product which translates audio data into coordinate data. Worked on the EchoNet Production Team to prepare the project for a production deployment and pitch the project to customers
- **EchoNet 2.0** – Architected the software design of EchoNet 2.0, and created the frontend and backend
- **Admin Panel** – Created a native admin panel to control starting/stopping EchoNet
- **ML/AI Pipeline** – Created a pipeline to automate going from raw data to data snipping, to a model, to testing. Collected and trained models using the pipeline
- **EchoMesh** – Created a distributed system of multiple edge devices
- **Presentations** – Gave a presentation for internal stakeholders, and a video presentation for the Army
Tools Used: Python, JavaScript, PyTorch, torchaudio, wandb.ai, FastAPI, sockets, tkinter, Node.js, Next.js, React.js, PostgreSQL, Nvidia Jetson Platform, Raspberry Pi Platform, Docker, Linux, Ubuntu, Debian

(Reston, VA) **The EchoNet Project (Internship) (May 2023–Aug. 2023):** (Project description above) Worked on Team Edge to integrate the models and frontend into a usable product
- **EchoNet Edge Device** – Designed the physical Edge device and its components; designed and implemented the backend using a microservices architecture
- **Drone Live Video Feed** – Made the drone video feed available in the Flutter UI though the HLS protocol
- **K-12 Outreach** – Helped with code demos and shared my story
Tools Used: Python, Flask, Simple Websocket, PyAudio, Nginx, PostgreSQL, Psycopg, PostgREST, RTSP, hls.js, Octo TotoCV, Docker, Podman, OpenShift, Linux, RHEL, Fedora, Debian, Bash, Powershell

Projects

Statistics Papers (Dec. 2021–May 2024): Explored trends through statistics papers
- *Who are the fastest typists? (2024)* – Collected data about college-aged students to compare current trends with what has historically been normal typing speed
- *Movie Trends Over Time: A Short History (2023)* – Analyzed historical trends in movies in the context of technology and made data driven predictions for future trends in filmmaking
- *Factors for Performance (2021)* – Analyzed factors that affect grades of K-12 Students
Tools Used: SAS, Markdown, LaTeX, Libreoffice Writer, Python, NumPy, Scikit-learn, Pandas, SciPy, Beautiful Soup, Matplotlib, R (Programming Language), R-Markdown

College Football Stats (Apr. 2024–Jan. 2025): Created a website to view CFB statistics
Tools Used: JavaScript, HTML, CSS, express.js, MongoDB

Home Lab (Sep. 2017–Oct. 2024): Managed a home server to automate encoding movies to the AV1 codec
Tools Used: Linux, Debian, Gentoo, FreeBSD, Bash, Powershell, rav1e (libaom-av1), libvp9, FFmpeg, Samba, NFS, Jellyfin, Emby, Plex, VLC, Docker, Ansible, Btrfs, ZFS

Cryptography (July 2023–Feb. 2024): Implemented encode and decode algorithms, and used frequency analysis to attack them
Tools Used: Haskell, Python, Jupyter, Scikit-learn, R (programming language), Matlab

Education

University of Maryland, College Park (2020–2024)

B.S. Mathematics

B.S. Computer Science

Chantilly High School Advanced Diploma (2016–2020)

Certifications

DoD Secret Clearance

ICAgile Foundation of Dev-Ops (ICP-FDO)

IC-Agile Certified Professional (ICP)

AWS Certified Cloud Practitioner (CCP)

Skills

English (Native)

Chinese (Conversational)

Numeric Analysis

Real Analysis

Probability Theory

Context-Free Grammar

Git Version Control

Dockerization

Organizations

The Mighty Sound of Maryland (MSOM)

TBΣ Honorary Sorority

XR (VR & AR) Club

Terrapin Teachers

Maryland Ski Team

Accessibility and Disability Services (ADS) Notetaker

First Robotics Competition (FRC)

The Mighty Marching Chargers (MMC)

Experimental Chinese School (ECS) TA