Sid Su

Machine Learning Engineer, Full Stack Developer, Mathematician

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

Work History

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

DoW DCSA PEO "Investigation Management (IVM)" Ramp Team Lead (Apr. 2025-PRESENT): Modernization and Migration to Cloud, Defense Counterintelligence and Security Agency. An effort to modernize the software that background investigators use when clearing individuals to work for the DoW.

Ramp Team Lead (Feb. 2026-PRESENT): As the team lead of the Ramp Team, managed 5 developers, surging up to 8 during the migration, to build out core features and generalize VIP to become the basis of the larger IVM platform (IVM 1.0.0)

- **Customer Facing Work:** Professionally worked with both the product owner and users of the VIP system to listen to and implement enhancements to the VIP system
- **New Features:** Led in the development of new features, such as enhanced ROI creation with Gotenberg, the substitute feature, form fragments, performance optimization
- **Team Lead:** Broke up features into units of work for the team, and gave expert opinions to merge requests, ensuring a high level of code quality
- **Migration Taskforce:** Managed a surge of 8 developers for the migration from VIP to IVM 1.0.0
- **Architecture:** Worked with the architecture team come up with the migration plan, and a more general architecture
- **Temporal:** Converted critical methods to workflows and activities, enabling durable execution
- **Kafka:** Used Kafka streams to optimize performance via materialized views in the microservices architecture
- **Agentic Coding:** Lead my team in fully adopting agentic coding and specification-driven development

VIP Prototype Lead (Nov. 2025-Feb. 2026): A 90-day idea-to-production pilot in the DoW DCSA cloud, enabling background investigators to conduct virtual interviews on continuous vetting (CV) financial alerts. Worked as the sole backend developer and team lead to implement the REST API

- **REST API:** Developed a REST API to provide users with modern capabilities to resolve alerts
- **CI/CD:** Developed Docker images, and created a pipeline to deploy them automatically on changes
- **Adapters:** Wrote a set of adapters from government and the major 3 credit bureaus' data into the VIP
- **Adapter Taskforce:** Led 5 developers on loan from the Application team to create the adapters
- **ROI Creation:** Automated the previously manual process of generating an output PDF for adjudication
- **Data Modelling:** Worked with the Data team to create an SQL schema – now the core of the IVM project
- **Testing:** Worked with the testing team to recreate and fix bugs

Application Team Intermediate Backend Developer (Apr. 2025-Nov. 2025): As a backend developer on the Application team to create the platform which IVM sits on

- **REST API:** As a team created the standard structure of all Spring projects on IVM
- **Spring and Java Best Practices Document:** Created a best practices guide for the project
- **Business Rules Engine Evaluation:** Evaluated IBM ODM as a potential rules engine
- **Preliminary Data Modelling:** Worked with the data team to create a preliminary data model

Languages Used: Java, SQL, HTML, CSS, Typescript, Javascript
Libraries Used: Spring, Spring Boot, Thymeleaf, Angular, Apache PDFBox, H2 Database
Services Used: Temporal, Kafka, Kogito, Gotenberg, PostgreSQL, H2, PostgreSQL, Oracle RDBMS, IBM Operational Decisions Manager (ODM)
Tools Used: Postman, Insomnia, Insomnium, Alpine Linux, wkhtmltopdf, AWS Route 53, AWS ALB, AWS ECS, AWS ECR, GitLab CI, Docker, OpenCode

DoD CDAO "Project Harbinger" (Jan. 2025-Apr. 2025) (IBM Highest Metrics): A DoD Secret machine learning image segmentation project. Worked as a Machine Learning Engineer focusing on the preprocessing and postprocessing of data

- **Continuous Inferencing Improvements:** Reworked the continuous inferencing strategy, improving live model performance
- **Novel Preprocessing Methods:** Implemented preprocessing methods to improve detection of a class
- **Novel Postprocessing Methods:** Implemented postprocessing methods for smoother output
- **Novel Machine Learning Approaches:** Implemented recent novel machine learning approaches

Tools Used: Python, PyTorch, Scikit-Learn, Pandas, Numpy, MySQL, Bash, Makefile, Docker, EMACS

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

Education

University of Maryland, College Park (2020-2024)

B.S. Mathematics

B.S. Computer Science

Chantilly High School (2016-2020)

Certifications

DoW Top Secret Clearance

CompTIA Security+

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

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

The EchoNet Prod Team (June 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

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

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

Semantic Logic Papers (Jan. 2021-Jan. 2022): Wrote semantic philosophy papers:

- **The Theory of Conditionals (2021):** Argued for the NTV theory of Indicative Conditionals
- **Causation in Counterfactuals (2021):** Argued for the Causal Model of Counterfactual Conditionals
- **The Problem of Pronouns (2022):** Explored the pronoun binding problem in Variable-Free Semantics

Tools Used: Haskell, Markdown, LaTeX

College Football Stats (Apr. 2024-Jan. 2025): Created a website to view CFB statistics

Tools Used: JavaScript, HTML, CSS, express.js, MongoDB

Home Lab (Sept. 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