SANCHIT SAHAY

New York University, New York, NY ◇ ss19723@nyu.edu ◇ stupendoussuperpowers.github.io

SUMMARY

Software engineer with 3+ years of experience building production-scale distributed systems and contributing to systems research in operating systems and software supply-chain security with a strong background in cloud and virtualization.

EDUCATION

New York University, Tandon School of Engineering, New York, NY — Sept 2024 – May 2026
Master of Science, Computer Science — GPA: 3.889/4.0

Manipal Institute of Technology, Manipal, India — June 2018 – July 2022
Bachelor of Technology - Information Technology (Minor: Big Data Analysis) — GPA: 9.10/10

RESEARCH AND WORK EXPERIENCE

Secure Systems Lab, NYU, New York, NY — June 2025 – Present
Researcher, Advisor: Prof. Justin Cappos — *Rust, Go, C, eBPF, Linux Kernel & FS*

Lind-Wasm: *WebAssembly-based sandbox enabling secure, memory-safe execution of POSIX applications through a lightweight kernel microvisor.*

- Designed and implemented syscall monitor APIs enabling fine-grained interception and override of system calls issued by sandboxed processes, providing observability and runtime control over execution behavior.
- Architected and built the in-memory filesystem, enabling trusted I/O for secure C compilation inside Intel SGX enclaves.
- Developed a comprehensive benchmark suite to quantify isolation overheads, syscall interposition costs, and filesystem throughput, informing optimization decisions for syscall resolution mechanisms.
- Extended the Lind runtime to support compilation and execution of Rust programs.

SBOMit: *OpenSSF initiative to augment Software Bills of Materials with build time in-toto attestations.*

- Extended in-toto with eBPF-based network attestations to record package sources and detect suspicious network calls during builds, exposing attack vectors invisible to manifest-based SBOM tools.
- Integrated attestation support into Syft, demonstrating concrete blind spots in widely-used SBOM generators.
- Presented implementation and findings at **KubeCon + CloudNativeCon, Atlanta** (Nov 2025); accepted speaker at **PyCon US, Long Beach** (May 2026).

Commvault Systems, Bangalore, India — Jan 2022 – Aug 2024
Engineer, Virtual Server Agent Team — *Python, .NET, VMware, Huawei Cloud*

- Built data-protection software for VMware and Huawei private-cloud environments, deployed across 500+ enterprise and government organizations.
- Developed Commvault's VMware Cloud Director plugin, streamlining multi-step data-protection workflows and reducing operational overhead for managed-service customers.
- Refactored the Python SDK for VMware Cloud Director, improving end-to-end test reliability.
- Implemented agent-based Guest OS file recovery for VMware VMs to enable deduplicated source-to-storage communication, eliminating bottlenecks for workloads exceeding 10 GiB or 10,000 files.
- Extended Application-Aware Backup support to VMware Cloud Director VMs, enabling granular backup and recovery for SQL Server and other workloads.

LegalAI, Remote Apr 2021 – Dec 2021
Full-Stack & DevOps Intern *Node.js, React, Google Cloud Platform*

- Designed and built an end-to-end claims-processing platform composed of multiple GCP-hosted microservices and React portals, handling legal-draft generation and review.
- Built the CI/CD pipeline with a custom local App Engine–like runtime to emulate GCP behavior, ensuring parity between local development and cloud production environments.

OPEN SOURCE AND COURSE PROJECTS

HFS+ Port For FreeBSD: C FreeBSD Status Report

- Ported Apple's open-source HFS+ filesystem to FreeBSD 14, adapting kernel VFS-layer operations to modern FreeBSD interfaces.
- Developed userland utilities for mounting and management of HFS+ volumes.

Cargo: Rust github.com/stupendoussuperpowers/cargo

- Contributed to Cargo, Rust's official package manager and build system.

Improving Learned Bloom Filters github.com/stupendoussuperpowers/wise-bloom-filters

- Benchmarked and compared optimization techniques for Learned Bloom Filters. Evaluated trade-offs between false-positive rate and memory efficiency for Projection Hashing, Caching, and LoRA.

Talk2Data: Python, Google Cloud Platform (GCP) github.com/Sitanshuk/Talk2Doc

- Built scalable data pipelines on GCP to extract, organize, and serve email and Notion data for student productivity including job applications, course materials, and deadlines.
- Implemented RAG-powered queryable interfaces and chatbots using personalized LLMs, with load-efficient hosting on Google Cloud Platform.

MTA Ridership Prediction: Python github.com/stupendoussuperpowers/mta-ridership

- Trained ML models to predict NYC subway ridership from temporal and fare-class features; applied K-Shape clustering to uncover neighborhood-level ridership patterns across the system.

TECHNICAL SKILLS

Languages: Rust, Python, Go, C/C++, Node.js, Java, C#
Frameworks: eBPF, React, Next.js, .NET, Android SDK
Databases: MongoDB, SQL Server, PostgreSQL, Redis
Cloud & Virtualization: GCP, AWS, VMware vCenter and Cloud Director, Huawei FusionCompute, Docker, Kubernetes
ML/Big Data: Pandas, NumPy, Keras, Spark, Hadoop, neo4j, CUDA

RELEVANT COURSEWORK

New York University
Cloud Computing and Big Data, Software Supply Chain Security, High Performance Machine Learning, Algorithmic Machine Learning and Data Science, Computer Networking

Manipal Institute of Technology
Operating Systems, Database Management Systems, Distributed Systems, Software Reliability