

Daniel David

Software Engineer | Specialized in Scalable & Secure ML Pipelines | Columbia University
New York, NY 10025 | (202) 594 4804 | daniel1957000@gmail.com | linkedin/in/ddavid37 | github.com/ddavid37

WORK EXPERIENCE

Software Engineer - Distributed ML Systems – Federated Learning

Dec 24' – Nov 25'

Rhino Federated Computing - [Link](#)

- **Distributed Workflow Engineering:** Engineered distributed ML workflows using NVIDIA NVFlare, refactoring Client APIs to optimize computing efficiency and inter-node communication for PyTorch models. [Link](#).
- **System Reliability & Monitoring:** Implemented expanded network topology support and real-time system monitoring using Grafana, allowing for the diagnosis of latency bottlenecks in distributed learning sessions.
- **Infrastructure Optimization:** Migrated legacy federated learning examples to modern architectures by optimizing Dockerfiles and configurations, streamlining the deployment across heterogeneous compute environments.
- **Pipeline Scaling:** Supported the scaling of ML pipelines for biotech partners (Eli Lilly), facilitating the training of high-value models (TuneLab) on distributed datasets by ensuring robust container orchestration. [Link](#).
- **Encryption & Data Integrity Mechanisms:** Architected secure model transmission protocols leveraging AWS Secrets Manager, ensuring data integrity during multi-node parameter aggregation cycles. [Link](#).
- **Documentation & Onboarding:** Authored technical documentation regarding SDK usage and platform guides, significantly reducing onboarding friction, in addition, provided technical support throughout the process. [Link](#), [Link](#).

TECHNICAL KNOWLEDGE

- **Languages:** Python (Advanced), C, C++, Java, SQL, Assembly.
- **Machine Learning & AI:** Core industry expertise in Machine Learning and Deep Learning from Rhino with PyTorch, TensorFlow, scikit-learn, NumPy, Pandas. Built and deployed different models on distributed cloud systems. Experience with LLMs, RAG pipelines and agentic workflows using frameworks like LangChain and internet retrieval tools like Tavily.
- **Federated Learning & Security:** Core professional expertise from Rhino Federated Computing based on NVIDIA's NVFLARE framework. Authored onboarding tutorials and developed AES+RSA encryption workflows (integrated with AWS Secrets Manager and manual key handling) aligned with enterprise and research-grade standards. [Link](#).
- **Systems & Infrastructure:** Experienced with distributed computing across multi-cloud environments, CI/CD pipelines, virtualization. At Rhino, consistently used Grafana for cloud monitoring and real-time visibility for client services.
- **Cloud & DevOps:** Extensive professional experience with containerization (Docker) and orchestration (Kubernetes), deploying federated workloads across AWS, GCP, and Azure.
- **Databases & Data Engineering:** Proficient in SQL and PostgreSQL, with deep experience at Rhino configuring schemas, validating federated datasets, and supporting production data pipelines across healthcare and finance clients. Worked extensively with cloud database integration (AWS/GCP/Azure) to enable scalable federated workflows. Skilled in query design and schema optimization, with additional academic exposure to NoSQL systems (MongoDB, Neo4j).
- **Security & Systems:** Experience with Network Security, Static/Dynamic Analysis, Memory Management, Reverse Engineering tools (IDA Pro, Wireshark, Procmon).
- **Web & App Development:** Proficient in JavaScript, CSS, React, and Bootstrap for web apps. Built interactive ML learning platforms using Flask + JS + Bootstrap with interactive quizzes, progress tracking, and responsive UI. [Link](#).

EDUCATION

Bachelor of Science - Computer Science

Overall GPA: 3.81/4.00, ML/AI track, Recipient of the "Renaissance" Honors Scholarship, Dean's List, PTK Honor Society.

SELECTED PROJECT

Tutorial #1 – Basic Usage of Rhino FCP for Federated Learning [Link](#)

Created the flagship tutorial for the Rhino FCP to guide users through core fundamental steps of the platform. Key steps include setting up projects, importing datasets from cloud-mounted storage, and running containerized codes for training, inference, etc.

Model & Code Encryption for Secure Federated Computing Workflows [Link](#)

Developed a workflow to encrypt ML models and training logic in Rhino's Federated Computing Platform, with secure key handling via AWS Secrets Manager or JSON keyfiles along with article that now serves as an advance reference guide.

Linear Algebraic Approaches to Neuroimaging Data Compression [Link](#)

Comparative analysis of Matrix (SVD) and Tensor (Tucker) Decomposition methods for compressing high-dimensional data. Demonstrated that Tucker decomposition offers superior reconstruction fidelity and perceptual similarity by preserving multi-dimensional structural relationships, contrasting it with SVD's performance in extreme compression scenarios.

Personal Interests: Innovation, Geography, Long-Running and Endurance Sports, Skiing, Basketball.