

Van Minh Nguyen

PH.D. CANDIDATE · MLOPS ENTHUSIAST

Melbourne, FL USA

☎ (+1) 321-368-9310 | ✉ vmnguyen251@gmail.com | 🏠 <https://n0k0m3.github.io/> | 📷 n0k0m3 | 🌐 minhnguyen251

Education

Florida Tech

Melbourne, FL

PH.D. OPERATIONS RESEARCH

Aug 2020 - Present

- Determine bacteria mutation rate with double stochastic branching process with random offspring
- Researching privacy focused, temporal synthetic Electronics Health Records (EHR) generation with Differential Privacy.

M.S. OPERATIONS RESEARCH, GPA: 4.0

Aug 2018 - May 2020

B.S. BIOCHEMISTRY (BIOLOGY EMPHASIS), GPA: 3.43

Aug 2014 - May 2018

Work Experience

Florida Tech

Melbourne, FL

DEVOPS ENGINEER - NEURAL TRANSMISSIONS LAB

Jan 2022 - Present

- Deploy and maintain on-premise bare-metal servers for research lab.
- Deploy multi-user research/development environments with GPU support for deep learning.
- Deploy Zero Trust Architecture for granting role-based access (RBAC) to servers: authentication with Keycloak and SAML, authorization with Google OAuth and OpenLDAP.
- Secure deployments with HTTPS for web server, with proper DNS configuration, short-lived SSH certs for remote access, and VNC over HTTPS for remote control.

TEACHING ASSISTANT - DEPARTMENT OF MATHEMATICAL SCIENCES

Aug 2018 - Dec 2022

- Taught and graded exams for Probability & Statistics, Neural Networks, Calculus I, II, III.
- Aided students studying Stochastic Modeling and Theory of Stochastic Processes.
- Technical support for students learning Neural Networks and Machine Learning.

Truveta

Seattle, WA

RESEARCH INTERN

Jan 2022 - May 2022

- Cost-effective and scalable SOTA Named Entity Recognition (NER) pipelines for clinical notes information extraction and de-identification. Saved \$2 million in annual cost, and further optimized to reduce 75% operating cost overhead compared with original baseline model.
- Prototyped clinical notes annotation tool for internal annotator team.
- Threat modeling and mitigation for deployment of mentioned tools, including common attack vectors on platform and PHI leakage threats.

GRADUATE INTERN

May 2021 - Aug 2021

- Developed an ETL pipeline to measure data quality of Truveta Health Data Model (THDM), garnering trusts from health providers. Contributed to new partnerships with multiple after Private Review in Aug 2021, and with Microsoft in Sept 2021.
- Designed a synthetic patient data model for stress-testing and identify bottleneck in ETL process.
- Developed an annotation recommender system for medical concept normalization. Reduced internal annotators' workload by 80%

Skills

Software & Tools

TensorFlow, PyTorch, ONNX, Docker, Kubernetes, Spark/PySpark, Hadoop Streaming, SQL/MariaDB, NoSQL/Redis

Operating Systems

Ubuntu Server, Arch Linux

Cloud Infrastructure

Microsoft Azure, Google Cloud Platform, Databricks, Kubeflow, MLFlow

Other Skills

Data Processing & Analysis, Stochastic Modeling, Reverse Engineering, Penetration Testing

Projects

3D Reconstruction of satellite using Dynamic Neural Radiance Fields

Compare *instant-ngp* and *D-NeRF* performance in 3D model reconstruction of satellite from a single *stationary* camera video of the real object.

Temporal-Spatial Transformer in Soft Actor-Critic/TD3 for autonomous driving

[\[Project link\]](#)

Implement temporal-spatial transformer module with action memory in Soft Actor-Critic/TD3 architecture in "highway-env", an OpenAI Gym compatible environment for autonomous driving decision-making tasks.

Mobile Game Assets Decryption and Datamining tool

[\[Repo link\]](#) [\[Data link\]](#)

Reverse Engineered game using Frida (dynamic analysis) and Ghidra (static analysis) for assets decryption for game "Date A Live: Spirit Pledge". Scheduled cron Pipeline for continuous update, decrypt, and datamine.

GPU-supported PySpark Notebook with DeltaLake

[\[Repo link\]](#)

Docker container for data analysis with Jupyter notebook server, RAPIDS AI, PySpark for GPU-accelerated, distributed and scalable ETL, with addition of all common data science libraries. The project aims for feature parity with Databricks.

Persistent Homology (TDA) feature engineering on Handwritten Digits and Letters

[\[Project link\]](#)

Apply Persistent Homology, a topological data analysis (TDA) technique for feature engineering on extended MNIST. 91% testing accuracy with a non-convolutional feedforward neural network.