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 - Dec 2023*

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

M.S. Operations Research — *Aug 2018 - May 2020*

B.S. Biochemistry (Biology Emphasis) — *Aug 2014 - May 2018*

Work Experience

Florida Tech — *Melbourne, FL*

Suicide Prevention Research - Department of Computer Engineering and Sciences — *Jan 2023 - Present*

- Developed a pipeline to scrape Twitter and Reddit data for research
- Preprocessed collected data and ingested into MongoDB
- Prototyped a simple NLP model (CBOW-based) to predict suicidal tendency from social media posts

MLOps Technician - Neural Transmissions Lab — *Jan 2022 - Present*

- Deploy and maintain on-premise bare-metal servers for research lab.
- Deploy multi-user research/development environments with role-based access control (RBAC) and GPU support
- Secure deployments with HTTPS, short-lived SSH, and VNC over HTTPS for remote control.

Teaching Assistant - Department of Mathematical Sciences — *Aug 2018 - Dec 2022*

- Taught and graded exams for Probability & Statistics, Calculus I, II, III. Tutored for Stochastic Modeling and Theory of Stochastic Processes.
- Technical assistance for students learning Neural Networks and Intro to Machine Learning.

Truveta — *Seattle, WA*

Research Intern — *Jan 2022 - May 2022*

- Researched and deployed scalable Named Entity Recognition (NER) pipelines for clinical notes information extraction and de-identification. Saved $2 million in annual cost, reduced 75% operating cost overhead compared with third party solutions and original baseline model.
- Prototyped clinical notes annotation tool for internal use.

Graduate Intern — *May 2021 - Aug 2021*

- Developed an ETL pipeline to measure data completeness, conformance, and plausibility of Truveta Health Data Model (THDM).
- Designed a synthetic patient data model for stress-testing and identify bottleneck in ETL process.

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

Uses *instant-ngp* and *D-NeRF* for 3D model reconstruction of satellite from a single view camera video of the real satellite object. Accepted at 33rd AAS/AIAA Space Flight Mechanics Meeting

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.

Persistent Homology 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.

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. The project aims for feature parity with Databricks.