

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.
 - Publication: Determination of Mutation Rates with Two Symmetric and Asymmetric Mutation Types. Symmetry. 2022; 14(8):1701.
- 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 (tweets) and Reddit data (posts and comments) for research
- Preprocessed millions of collected text 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

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

TEACHING ASSISTANT - DEPARTMENT OF MATHEMATICAL SCIENCES

Aug 2018 - Present

- 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 pipelines for clinical notes information extraction and de-identification. The developed pipeline and model saved \$2 million in annual cost, reduced 75% operating cost overhead compared with third party solutions and baseline model.
- Prototyped clinical notes annotation tool for internal use.

GRADUATE INTERN

May 2021 - Aug 2021

- Developed an ETL pipeline to measure data quality of Truveta Health Data Model, which is now part of Truveta Studio, the first Truveta's product.
- Designed a synthetic patient data model for stress-testing and identify bottleneck in ETL process.

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 speech 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.

Skills and Interests

Skills

Python, R, C#; TensorFlow, PyTorch, ONNX, Docker, Kubernetes, Spark/PySpark, Hadoop Streaming, SQL/MariaDB, NoSQL/Redis; Microsoft Azure, Azure DevOps (CD/CI), Google Cloud Platform, Github Action (CD/CI), Databricks, Kubeflow, MFlow, Cloudflare Zero Trust; Data Mining, Data Processing & Analysis, Statistical Modeling, Stochastic Modeling, Mathematical Analysis

Interests

Reverse Engineering (Dynamic and Static Analysis), Self-hosting Homelab and Home automation, Soccer, Badminton, PC Building