Van Minh **Nguyen**

Melbourne, FL USA

(+1) 321-368-9310 | vmnguyen251@gmail.com | https://n0k0m3.github.io/ | n0k0m3 | minhnguyen251

Work Experience

Truveta — *Seattle, WA*

Research Intern — *Jan 2022 - May 2022*

- Researched, trained, and deployed scalable NER pipelines for clinical notes information extraction and de-identification using SparkNLP with PyTorch API. The pipelines saved $2 million in annual cost comparing with third party solutions, reduced 75% operating cost comparing with baseline model.
- Threat modeled using OWASP Threat Dragon and suggested mitigation for deployment of aforementioned pipelines and tools.
- Prototyped clinical notes annotation tool for internal use based on Label Studio and INCEpTION.

Graduate Intern — *May 2021 - Aug 2021*

- Developed an ETL pipeline to measure data quality (completeness, conformance, and plausibility) of Truveta Health Data Model using Spark and Azure Pipelines, contributed to new partnership with Microsoft in Sept 2021. The pipeline is now part of the first Truveta's product - Truveta Studio.
- Designed a synthetic patient data model for stress-testing and identify bottleneck in ETL process, generating millions of synthetic records in 1 hour.
- Developed an annotation recommender system for medical concept normalization. Reduced internal annotators' workload by 80%
- Assisted team members and other co-workers on various Projects and Collaborative efforts.

Florida Tech — *Melbourne, FL*

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

- Developed improved pipelines to scrape Twitter (tweets) and Reddit data (posts), reduced ingestion time by 60 times using pure bash
- Prototyped a simple model to predict suicidal tendency from social media posts, using NLP features like text statistics and TF-IDF
- Analyzed changes in monthly word statistics and wordcloud of suicidal posts over 5 years period

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

- Deployed and maintained on-premise bare-metal servers for research lab (Kubernetes cluster on Ubuntu Server using Kubeadm).
- Deployed multi-user research/development environments (Jupyterhub on Kubernetes) with GPU support and role-based access control (Keycloak)
- Secured exposed deployments with HTTPS and proper DNS configuration, 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.

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*

Projects

3D Reconstruction of satellite using Dynamic Neural Radiance Fields — *[Publication]*

Uses *instant-ngp* and *D-NeRF* for 3D model reconstruction of satellite from a single view camera video of the real satellite object.

Temporal-Spatial Transformer in Soft Actor-Critic/TD3 for autonomous driving — *[Project link]*

Implement temporal-spatial transformer module in *PyTorch* 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, using *giotto-ai*, on extended MNIST (MNIST with additional handwritten letters) dataset. 91% testing accuracy with a non-convolutional feedforward neural network using *Keras/TensorFlow*.

Sentiment Analysis on MyAnimeList User Ratings — *[Project link]*

Datamined and created pipeline to extract user review from MAL using *Redis* cache and REST API. Predict user rating based on data-mined review text with RNN using *Keras/TensorFlow*. 94% testing accuracy with ensemble model.

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, MLFlow, 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