

Van Minh Nguyen

Melbourne, FL USA

✉ 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
- Research privacy-focused, longitudinal (temporal) generation of synthetic Electronic Health Records (EHR) with Differential Privacy

M.S. OPERATIONS RESEARCH, GPA: 4.0

Aug 2018 - May 2020

B.S. BIOCHEMISTRY (BIOLOGY EMPHASIS), GPA: 3.6

Aug 2014 - May 2018

Work Experience

Truveta

Seattle, WA

RESEARCH INTERN

Jan 2022 - May 2022

- Developed and deployed scalable Named Entity Recognition (NER) pipelines for clinical notes information extraction and de-identification using John Snow Labs' SparkNLP free model and a custom PyTorch model
- Analyzed cloud infrastructure costs and optimized potential savings by calculating cluster size, resulting in \$2M annual savings and a 75% reduction in operating costs
- Conducted threat modeling using OWASP Threat Dragon and recommended mitigation strategies for pipeline deployment
- Created a clinical notes annotation tool prototype based on Label Studio and INCEpTION for internal use
- Navigated complex organizational processes and collaborated with multiple stakeholders to ensure timely access to resources and maintain project momentum

GRADUATE INTERN

May 2021 - Aug 2021

- Developed a data quality measurement toolkit using Spark for the Truveta Health Data Model (THDM), which assessed data completeness, conformance to THDM, and plausibility based on OHDSI criteria
- The toolkit garnered trust from participating health providers, leading to multiple partnerships with new healthcare providers, a collaboration with Microsoft, and integration into Truveta Studio, Truveta's first product
- Established the foundational approach and model for Synthetic Patient Health Data, leveraging probabilistic theory of document retrieval, representation learning, and deep learning, with applications in generating synthetic health data, filling missing patient data, and predicting doctor-patient encounters and diagnoses/medications from health history
- Designed and implemented a synthetic patient data model using Monte Carlo sampling for stress-testing and bottleneck identification in the ETL process, capable of generating millions of records in 1 hour, ensuring robust data ingestion and querying systems
- Developed an annotation recommender system for medical concept normalization, reducing annotators' workload by 80%
- Deployed and integrated the data quality measurement toolkit into the data ingestion pipeline using Azure DevOps Pipelines, leveraging Azure resources such as Databricks, Azure VM, Storage Blob, and Data Factory (ADF)

Engage-AI.org

Remote

DATA ENGINEER - CONTRACTOR

May 2023 - Jan 2024

- Spearheaded the initial development and deployment of the Engage AI Data Platform, leveraging Cloudflare R2 for storage, DuckDB for query processing, and Apache Superset for data visualization
- Achieved a zero-cost initial deployment and laid the groundwork for a cost-efficient, cloud-based data management solution, setting the foundation for Engage AI's data platform
- Collaborated with data analysts to understand their data requirements, refining and optimizing the platform based on feedback

Florida Tech

Melbourne, FL

MLOPS TECHNICIAN - NEURAL TRANSMISSIONS LAB

Jan 2022 - Present

- As the sole architect and manager, initiated and implemented an on-premise server cluster for the research lab, using Kubernetes on Ubuntu Server.
- Set up multi-user research environments with GPU support and role-based access control using JupyterHub, Kubernetes, and Keycloak
- Secured deployments using HTTPS, DNS configuration, short-lived SSH, and VNC over HTTPS

TEACHING ASSISTANT - DEPARTMENT OF MATHEMATICAL SCIENCES

Aug 2018 - Present

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

IBM Skills Academy

Remote

DATA SCIENCE PRACTITIONER - INSTRUCTOR

May 2021 - Jun 2021

- Audited IBM Data Science learning material and proposed changes that was added to the material
- Created a curriculum for teaching IBM Data Science course at Florida Tech
- Earned Certification for Instructor role for IBM Data Science Practitioner

Research Experience

Florida Tech

SUICIDE PREVENTION RESEARCH - DEPARTMENT OF COMPUTER ENGINEERING AND SCIENCES

Melbourne, FL

Jan 2023 - Jan 2024

- Enhanced data scraping pipelines for Twitter and Reddit, reducing ingestion time by 60 times
- Developed a prototype model for predicting suicidal tendencies from social media posts using NLP features, achieving an accuracy rate of 62%
- Analyzed monthly word statistics and word clouds of 800,000 suicidal posts over a 5-year period, providing critical insights for suicide prevention efforts

Projects

3D Reconstruction of satellite using Dynamic Neural Radiance Fields

[Publication]

Applied *instant-ngp* and *D-NeRF* for efficient 3D model reconstruction of satellite from a single view camera video of the real satellite object, enabling improved space debris removal and on-orbit servicing

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

[Project link]

Implemented a transformer module and action-memory within Soft Actor-Critic and TD3 architectures, enabling the agent to "remember" its previous actions and effectively predict the next ones. Tested on *highway-env*, an OpenAI Gym environment for autonomous driving decision-making tasks

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, aiming for feature parity with Databricks - a popular cloud-based data analytics platform

Persistent Homology feature engineering on Handwritten Digits and Letters

[Project link]

Apply Persistent Homology, a topological data analysis technique for feature engineering, using *giotto-ai*, on extended MNIST dataset, achieving 91% testing accuracy with a non-convolutional feedforward neural network in *Keras/TensorFlow*

Skills

Programming & Deep Learning

Python, R, C#, TensorFlow, PyTorch, ONNX

Big Data & Cloud Platforms

Spark/PySpark, Hadoop Streaming, Microsoft Azure, Databricks, Kubeflow, MLFlow

Deployment & Databases

Docker, Kubernetes, Azure Pipelines, Cloudflare Zero Trust, SQL (MariaDB), NoSQL (MongoDB, Redis)

Analytics & Modeling

Data Mining, Data Processing & Analysis, Statistical Modeling, Stochastic Modeling, Mathematical Analysis

Publications

JOURNAL ARTICLES

Characterizing Satellite Geometry via Accelerated 3D Gaussian Splatting

Van Minh Nguyen, Emma Sandidge, Trupti Mahendrakar, Ryan T. White

Aerospace 11.3 (2024). 2024

promSEMBLE: Hard Pattern Mining and Ensemble Learning for Detecting DNA Promoter Sequences

Bindi M. Nagda, Van Minh Nguyen, Ryan T. White

IEEE/ACM Transactions on Computational Biology and Bioinformatics 21.1 (2023) pp. 208–214. 2023

Examining the Potential of Generative Language Models for Aviation Safety Analysis: Case Study and Insights Using the Aviation Safety Reporting System (ASRS)

Archana Tikayat Ray, Anirudh Prabhakara Bhat, Ryan T. White, Van Minh Nguyen, Olivia J. Pinon Fischer, Dimitri N. Mavris

Aerospace 10.9 (2023). 2023

Determination of Mutation Rates with Two Symmetric and Asymmetric Mutation Types

Jewgeni H. Dshalalow, Van Minh Nguyen, Richard R. Sinden, Ryan T. White

Symmetry 14.8 (2022). 2022

CONFERENCE PROCEEDINGS

Natural Language Explanations for Suicide Risk Classification Using Large Language Models

William Stern, Seng Jhing Goh, Nasheen Nur, Patrick J Aragon, Thomas Mercer, Siddhartha Bhattacharyya, Chiradeep Sen, Van Minh Nguyen

Machine Learning for Cognitive and Mental Health Workshop (ML4CMH 2024) at AAAI 2024, 2024

3D Reconstruction of Non-Cooperative Resident Space Object using Instant NeRF and D-NeRF

Basilio Caruso, Trupti Mahendrakar, Van Minh Nguyen, Ryan T. White, Todd Steffen

33rd AAS/AIAA Space Flight Mechanics Meeting 33.417 (2023). 2023

Conceptualizing Suicidal Behavior: Utilizing Explanations of Predicted Outcomes to Analyze Longitudinal Social Media Data

Van Minh Nguyen, Nasheen Nur, William Stern, Thomas Mercer, Chiradeep Sen, Siddhartha Bhattacharyya, Victor Tumbiolo, Seng Jhing Goh

2023 22nd IEEE International Conference on Machine Learning and Applications (ICMLA), 2023