Tung Thanh Le

Website: http://ttungl.github.io/ Mobile Phone: 612-490-3605
U.S. Permanent Residency Personal Email: ttungl@gmail.com

❖ ***Education***

- **University of Louisiana at Lafayette, USA**
  Doctor of Philosophy (Ph.D.) in Computer Science
  08/2013 – 12/2018
- **Kumoh National Institute of Technology, South Korea**
  Master of Engineering (M.Eng.) in IT Convergence Engineering
  09/2011 – 08/2013
- **University of Louisiana at Lafayette, USA**
  Master of Science (M.Sc.) in Computer Science
  08/2013 – 12/2016
- **Danang University of Technology, Vietnam**
  Bachelor of Engineering (B.Eng.) in Electrical Engineering,
  08/2002 – 08/2007

❖ ***Professional Work Experience***

❖ ***Interest***

- Algorithmic Optimization, Mathematical Modeling, BigData
- Machine Learning, Deep Learning and Artificial Intelligence

- **Lead Machine Learning Ops Engineer - *Thomson Reuters* -** *5/2023 – Present*
  - Developed an end-to-end solution to address customers' challenges in reviewing tax datasets by leveraging large language model (LLMs) to answer specific questions based on their tax data. Implemented and deployed the solution on Google Cloud, utilizing Gemini Pro with a function calling approach to convert questions to SQL queries and retrieve responses from Gemini, along with a user interface built using Streamlit. *Achieved the best working prototype Award* from *Corporate Tax &Trade Generative AI Hackathon*.
  - Developed an end-to-end solution to detect hallucinated AI-generated text using a Retrieval-Augmented Generation (RAG) approach, leveraging Azure for embedding models, Milvus vector database for similarity search to obtain relevant paragraphs, and OpenAI GPT-4 for generating answers and identifying potential hallucinations.
  - Developed ML DevOps features, deploying and maintaining ML services such as custom entity extractors service (NER), image processing, document processing.

- **Senior Data Scientist - *NBCUniversal* -** *12/2021 – 4/2023*
  - ***Lift Measurements:*** The goal is to measure the impact of advertising campaigns. Responsible for building ETL data pipelines with Python, PySpark, SQL on **Databricks** and **SnowPark** for data processing, feature engineering, feature selection, using matching methods such as propensity score matching for measuring the impact.
  - ***Face Recognition:*** The goal is to help data labeling on celebrity faces/brand objects in advertising video clips for conducting analysis on who contributed high sales/conversion rates in the advertising campaigns. Responsible for building the model using MTCNN, FaceNet, and SVM. MTCNN is used to capture facial areas from inputs. Faces captured are used for training FaceNet. SVM is used to classify new faces based on Face Embedding from trained FaceNet model. Implemented on **AWS EC2 Deep Learning instance**.

- **Data Scientist - *J.D. POWER* -** *07/2018 – 12/2021*
  - ***Days-to-turn on Vehicles Prediction****:* The goal is to help the OEM/dealers planning to optimally re-stock their sales inventories based on days-to-turn prediction. Responsible for building predictive models using data analytics, machine learning to predict days-to-turn target which determines how long it takes to sell a specific new car in the inventory. Implemented on **AWS** and **databricks** using Python and SQL, and **Tableau** and **Streamlit** for dashboards.
  - ***PIN Transformation****:* Building ETL big data pipelines from SAS to Python using BigQuery, PySpark, Python, Javascript for production on **AWS**, **GCP** platforms.
  - **Online Social Review Analytics:** The goal is to help evaluating the in-store performance rating based on the customers' reviews of the banks across U.S. Responsible for building the reviews sentiment analysis using natural language processing (**NLP**) techniques such as text cleaning, feature engineering using outlier remover, lemmatization, N-grams tokenization; Utilizing AWS Comprehend, SageMaker, Google Cloud NLP.

- **Research Intern - Hanwha Thales**, S. Korea - *08/30/2012 – 12/31/2012*
  Responsible for optimizing the network topologies for ships' built-in-network communication
- **Software Engineer - Unilab-DUT** (Novas Technologies Ltd.), Vietnam
  04/01/2008 – 06/01/2011: Responsible for software-hardware development.
- **Summer Intern – Orion Tech.**, S. Korea - *06/01/2012 – 08/30/2012*
  Responsible for programming network communication in ships.
- **PCB Layout& Design Engineer- Acronics Systems**, Inc -San Jose, CA
  06/01/2007 – 03/30/2008: Responsible for designing PCB on high-speed circuit boards.

❖ ***Projects***

- **Donation Analytics (Insight Data Engineering Challenge):** Analyzed loyalty trends in campaign contributions for cash-strapped political candidates by identifying zip codes with repeat donors and calculating their spending patterns.
- **Behavioral Cloning (Deep Learning):** Built and trained a convolutional neural network using TensorFlow, Keras, and Nvidia architecture for autonomous driving in a simulator. Performed image processing and augmentation with OpenCV. Utilized dropout, Adam optimizer, and Udacity dataset. Trained model on **AWS EC2**.
- **Network-on-Chip Optimization:** Designed the mathematical modeling for optimizing interconnections and energy efficiency in network-on-chip. Used **CPLEX**, **Gurobi** solvers, Python (**pyomo**), Matlab (heuristic algorithms), and machine learning algorithms for solving this optimization problem.

❖ ***Professional Certificates***

- Generative AI with Large Language Models (2023)
  Online Course – DeepLearning.AI
- Certification of Natural Language Processing Specialization (2021)
  Online Course – *DeepLearning.AI*
- Certification of Machine Learning (2017)
  Online Course – Stanford University
- Certification of Statistical Learning (2018)
  Online Course – Stanford University

❖ ***Honors & Awards***

- Graduate Teaching Assistantship, *09/2015 – 06/2018*
- NSF Graduate Research Fellowship, *09/2013 – 08/2015*
- Best Paper Award - 14th Conference on Electronics & Info. Communications *2012*
- NIPA scholarship and NRF scholarship, South Korea, *09/2011 – 06/2013*
- Samsung Thales scholarship for student travel in *12/2012*
- Excellent student, Danang University of Technology, *2004 –2007*
- One of four honor students achieving highest score on graduation thesis (4/500) in *2007*

❖ ***Computer Skills***

- **Programming languages:** Python, Java, PySpark, Scala, Rust, BigQuery, Javascripts, SQL, C/C++, R, MATLAB, CPLEX/AMPL.
- **Frameworks/Libraries:** Deep Graph Lib (Graph Neural Networks), Databricks, Airflow, Tensorflow, Keras, Apache Spark, Snowflake, Snowpark, MLLib, Node.js, OpenCV, Scikit learn, PyTorch, Spacy, nltk, OpenAI, AWS products, H2O.ai and driverless AI platform, Trax by Google.
- **Data Visualization:** Tableau, Power BI.
- **Cloud Services:** Amazon AWS, Google Cloud Platform, Azure Cloud.