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

❖ ***Professional Work Experience***

❖ ***Interest:*** Optimization, Math Modeling, ML, DL, AI, Data Science

- **Lead Artificial Intelligence SW Engineer – *Cast & Crew –*** *12/2024 – present*
  - ***Document Intelligence Extraction Processing:*** Built a **Generative AI**/**LLM** application to extract structured financial data from complex and unformatted Excel **spreadsheets** for payroll timecard generation in the entertainment industry. Leveraged advanced *prompt engineering* and *fine-tuning* techniques using **OpenAI a**nd **Claude** models to extract precise data, implemented with **Python, Langchain, LangGraph, RAG, FastAPI**, and deployed on **Azure** Cloud.
  - ***Tax Document Processing***: Developed an LLM-powered application to extract data from tax documents (PDFs and images) utilizing Python, OCR, and image processing techniques to ensure accurate and efficient data extraction.
  - ***Location Mapping API***: Integrated an Auto-Correction Mapping Location API using a three-tier validation system—exact matching, fuzzy matching, and LLM-powered fallback—to normalize location data while ensuring cost efficiency and high accuracy.

- **Lead ML Ops/Research Engineer – *Thomson Reuters –*** *5/2023 – 10/2024*
  - ***Ask Tax Talks:*** Built an AI-based (chat agent) end-to-end solution to address customers' challenges in reviewing tax datasets by leveraging **large language models** (**LLMs**) to answer specific questions based on their tax data. Implemented and deployed the solution on **Google Cloud**, utilizing Gemini **API call** with a **function calling** approach to trigger specific actions to SQL queries and retrieve responses, built user interface using Streamlit and **Python**.
  - ***DevOps:*** Developed **ML** features, deploying and maintaining **ML pipelines** for internal services using **Python** and **Rust**.

- **Senior Manager, 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 an end-to-end solution, from data collection, **image processing**, to build and train **deep neural net** models with MTCNN, FaceNet, and **supervised learning 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. Implemented **PyTorch** on **AWS EC2**.

- **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 EDA, **ensemble models** (i.e. LightGBM, XGBoost) with **time series** to predict days-to-turn target which determines how long it takes to sell a specific new car in the inventory. Implemented on **AWS**, **databricks** using Python, 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.

- **Software Engineer - Unilab-DUT** (Novas Technologies Ltd.), Vietnam
  04/01/2008 – 06/01/2011: Responsible for software-hardware development.
- **PCB Layout& Design Engineer- Acronics Systems**, Inc -San Jose, CA
  06/01/2007 – 03/30/2008: Responsible for designing PCB

❖ ***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**.
- **Advanced Lane Finding (Computer Vision):** Built an advanced lane-finding algorithm using distortion correction, image rectification, color transforms, and gradient thresholding. Identified lane curvature and vehicle displacement. Overcame environmental challenges such as shadows and pavement changes. Detected highway lane lines on a video stream. Used **OpenCV** image analysis techniques to identify lines, including Hough Transforms and Canny edge detection.
- **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.

❖ ***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), Langchain, LangGraph, RAG, FastAPI, 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.