Quan Duong

quan.duong@helsinki.fi (358) 417566525 ruathudo.github.io quan-duong-1a2247a9

PORTFOLIO: My recent projects are listed here https://ruathudo.github.io.

EDUCATION:

- **University of Helsinki** — *Phd candidate* (2022 - present)
  Computer Science - Generative AI for Computational Musicology
- **University of Helsinki** — *Master's Degree* (2017 - 2021)
  Computer Science - Algorithms and Machine Learning
- **Helsinki Metropolia UAS** — *Bachelor's Degree* (2012 - 2016)
  Information Technology
- **Hanoi University of Science and Technology** — *Bachelor's Degree* (2010 - 2012)
  Information Technology

ACADEMIA: My recent publications are here and also listed in google scholar.

- **Multimodal interactive system for music generation (on going)**
- **On Brazilian Drum Claves and Generating Rhythm Patterns out of Them**
- **An Unsupervised method for OCR Post-Correction and Spelling Normalisation for Finnish**
- **Benchmarks for Unsupervised Discourse Change Detection**
- **TFW2V: An Enhanced Document Similarity Method for the Morphologically Rich Finnish Language**
- **Program committee in NLP4DH workshop**

WORK EXPERIENCE:

Movial Oy — *Senior data scientist / engineer* (Sep 2021 - present)
- Develop architecture, ML models, and data analysis

University of Helsinki — *Research Assistant* (Jul 2020 - Dec 2020)
- Research on NLP - Seq2seq , text generation models

Frosmo Oy — *Machine learning engineer* (Jan 2018 - Jul 2020)
- Develop ML models, create ML pipeline, integrate to cloud services

Edvisto Oy — *Co-founder, full stack developer* (Jul 2015 - Dec 2017)
- Develop full stack web application - online video editor for education

Clipme Oy — *Full stack developer* (Jan 2015 - Jul 2015)
- Develop full stack ecommerce web application

Web2Fix Oy — *Web developer* (Jul 2013 - May 2014)
- Develop a custom CMS

Projects:

- **2024: Linux kernel vulnerability detection with ML -** Research on applying machine learning to kernel runtime verification events that are able to detect the possible vulnerability. (CNN-LSTM, GenAI, Synthetic data)
- **2023-2024: Offline AI assisted Chatbot -** Design and implement chatbot using the latest generative AI technology. The chatbot should work similarly like ChatGPT but with private data. The chatbot is able to answer questions based on provided knowledge and working without internet connection. Supports multiple languages and explores the knowledge of industrial sectors. (RAG, Layout parser, LLAMA, Langchain)
- **2023: eCommerce Data Platform -** Design and implement a data platform with management services for products and inventories from different data sources. (FastAPI, React, Azure, B2C Auth)

- **2022: Large scale machine learning architecture -** The ML pipeline is able to serve multiple models training and predicting. The pipeline is designed with automation in mind. The algorithms are used to automate warranty claim approval for one of the biggest truck manufacturers. (AWS pipelines, sagemaker, sklearn)
- **2022: Schedule optimization -** Implement algorithm to optimize scheduling for charging electric vehicles, power peak suppression and minimize the cost of charging for the station.
- **2021: Datalake -** Implemented data pipeline from client data sources to the datalake (AWS).
- **2021: New text similarity algorithm -** Proposed new algorithms which outperform previous methods on document similarity. Use an unsupervised ML method, so no annotation is needed from humans.
- **2020 - 2021: Pivot points detection for discourse change in time series text data -** Proposed a new method using deep learning and synthetic dataset as a framework for detecting the changing in the discourse of text over time. It was able to demo the events found with relative accuracy about the time on a synthetic dataset.
- **2020: Post OCR correction for Finnish -** The text from Optical Character Recognition (ORC) might have many errors due to image quality. Especially in Finnish, when a word is very long. I proposed a new method using the Transformers algorithm. Combine the context of the word to have the correct form of misspelled one. It is able to improve the quality of text for various OCR methods.
- **2020: Product Recommender System -** Implement ML recommender system to suggest similar products based on the images and categorical data. The products are in various commercial industries like Clothing, electronics, etc.
- **2019-2020: Article Recommender System -** Implement the recommender system to suggest semantically similar news articles. The articles are in 3 languages: Finnish, Swedish, and English. The system should be able to self maintain and require the least adjustments.
- **2018-2019: Non AI Recommendation system and Web components -** Worked on various projects with different kinds of products. Create a tracking for measuring the performance of recommendations. Do A/B testing. Implemented UI components that meet the client's requirements.
- **2018: VIP/Churn Prediction -** Implemented ML algorithm which returns the prediction for the user before they are actually VIP/Churn in the future based on user behaviors. Also, designed the pipeline to automatically get the data from client sources and output the prediction through API. (XGBoost, Flask)
- **2018: Smartwatch backend server -** Backend to sync information between smartwatch and mobile to cloud server.
- **2015 - 2017: Online video editing system -** Designed a distributed video processing architecture running on AWS services. It was able to process multiple videos and parallelly deliver the results to the end clients. The platform allows users to upload videos, images, and audio and edit directly on cloud environments.
- **2012-2015: Full-stack web development -** Custom CMS, commercial sites, etc.

AWARDS:

- **Won the First Prize of Helsinki Region Infoshare (HRI) for creating innovative applications based on open data.**
- **Deep learning certificate**
  https://drive.google.com/file/d/1VmBGxKlEEyr6YTSdONBq6QnrMbskVhkD/view?usp=sharing
- **Machine learning certificate**
  https://www.coursera.org/account/accomplishments/certificate/33XRJG7MPSFM
- **Eastern European Machine Learning Summer School certificate**
  https://drive.google.com/file/d/1y_Zsf_uBhPK2mgw6nYkR0-zBMIUnl5Gx/view

HIGHLIGHT SKILLS:

Machine Learning, Deep Learning, GenAI, NLP, LLM, VAE, GAN, Seq2Seq, Transformers, BERT.

OTHER SKILLS:

Project management, sales support, teaching and proposals writing.

TOOLS:

Python, Sklearn, Pytorch, Tensorflow, Langchain, LLM, FastText, Gensim, XGBoost, Pandas, Numpy, Django, FastAPI, Spark, SQL, NoSQL, React, AWS, Azure, Git, Linux, Docker.