


MSC. ANDREW VO

350137, Potong Pasir, Singapore ☎ (+65) 87306506 ✉ andrewvot16@gmail.com

EDUCATION

CLARK UNIVERSITY – MA, USA Master of Data science, online	2022 – 2024
BACH KHOA HCM UNIVERSITY – HCMC, VN Excellent Engineering Degree (PFIEV-French), Master's degree Equivalent	2014 – 2018
JAMES COOK UNIVERSITY – SINGAPORE English Proficiency Certificate, IELTS 6.5 equivalent	2022
HCM UNIVERSITY OF SCIENCE – HCMC, VN Machine Learning Advance certificate Big data (Spark) engineering certificate	2019

PROFESSIONAL EXPERIENCES (7 YEARS)

 EVONIK SEA – SINGAPORE (2.5 YEARS) <i>Senior Data Scientist, Team Leader, Onsite</i>	Nov 2021 – August 2024 <i>3 International Business Park, Singapore 609927</i>
❖ Project: AI Generative Project	
● Scope of work:	
✓ Crawling data from Wiki, Prnewswire, LinkedIn to create dataset	
✓ Clean data, create data pipelines to process data from raw to clean	
✓ Train, finetune LLM to make prediction model to predict circular plastic name, name matching	
✓ Rerank embedding search , combine fulltext search vs embedding search to bring better results	
✓ Benchmark and compare latest llm models	
✓ Evaluate Model accuracy using TopK, Precision-Recall-F1 Score, Reciprocal Ranking,...	
✓ Create project proposal to improve current project, doing product roadmap, resource management, package & price management	
✓ Lead team of 4 engineers to implement the project	
✓ Deploy models to aws, EC2, AWS Sagemaker, Azure Microsoft servers	
● Skills: AWS sagerMarker, aws s3, lambda, IAM, MLOps, ML pipelines. Docker, create machine learning architect app, Linux, Databrick, Lambda, API Gateway	
❖ Project: CV Scoring and Extraction Project	
● Scope of work:	
✓ Extract information in candidate resume, map skills, dob, exp,... into Evonik HR Database to store, filter, AI Matching, feed data into HRMS	
✓ Lead team of 6 Engineers to build backend side of platform to support HR Division to process recruitment more effectively	
✓ Deploy project to production, using AWS, EC2, Nginx, SSL	
✓ Finetune stable diffusion to generate CV background, create new cv using Evonik templates	
✓ Estimate & Optimize cost for using prompt of OpenAI (GPT3.5, GPT4.0)	
✓ Work with BA, Head of Development, stack holders to clarify the logics, give solutions	
● Skills: OpenAI prompt, estimate & optimize cost, AWS, Linux, docker, deployment	
❖ Achievement at EVONIK SEA:	
✓ Coach and train junior and fresher members from NUS to become competent AI scientist	
✓ Work with stack holders, data engineer team, other managers in Singapore, Germany, USA to get to know the logic, features, requirements, the scope of the projects. Come up with data insights from current data and plan for analyzing data in the future	
✓ Finetune and deploy name matching model, LLM NLP model, multi-tag classification model.	
✓ Leading mutil-projects in the same time, work with high pressure	



CARRO – SINGAPORE (2 YEARS)

Senior Data Scientist, Team Leader, Hybrid

Sep 2019 – Nov 2021

28 Sin Ming Ln, Singapore 573972

❖ **Project: Car Vision Health Estimation**

- **Scope of work:**

- ✓ Validate accuracy, QC of car-mask datasets which are labeled from freelancer team in Singapore, Indo, Vietnam
- ✓ Build the following models: car polygon detection, panel detection (license plates, rear, back-front, door,...), damage detection (scratch, reflection, dent, dirty) using Mask RCNN, Unet, Yolo, Detectron2
- ✓ Evaluate Computer Vision model accuracy, improvement, combine different models to create a final pipeline
- ✓ Write fast API, deploy project to AWS Sagemaker, EC2, S3, Lambda function

- **Skills:** AWS sagerMarker, aws s3, lambda, IAM, Docker, linux, deep learning algorithms

❖ **Project: Car Acoustic Health Estimation**

- **Scope of work:**

- ✓ Be a senior DS to implement the following models: Sound Event Detection, Car Make, Car Fired issues using EfficientNet - transfer learning, MFCC.
- ✓ Feature extraction, modeling with machine Learning classification with SVM, Random Forest, logistic regression, neuron network, ensemble model, Bayes model
- ✓ Write FastAPI, integrate API with Flutter mobile app

- **Skills:** Sound Event Detection, MFCC, EfficientNet, ML algorithms

❖ **Achievement at CARRO:**

- ✓ Work closely with BOD to build up the team up to 6 members in Computer Vision team, evaluate team's performance
- ✓ Lead the team to do computer vision projects follow Scrum daily meeting, weekly meeting report
- ✓ Design machine learning pipelines and deploy projects to AWS Sage Marker, S3, lambda functions. Apply deep learning to the models such as car detection, panel detection(objects of car), car damage detection, using Mask RCNN, Detectron2, transfer learning. Object detection using YoloV5, EfficientNet



GENESOLUTIONS – HCMC (2 YEARS)

Data Scientist, Team Leader

Jan 2018 – Nov 2019

186 Nguyen Duy Duong, HCMC, 700000

❖ **Project: Abnormal Variants Diagnose – Gene Product**

- **Scope of work:**

- ✓ Build and manage a team of 5 members to implement projects, follow scrum meeting
- ✓ Build data pipelines from raw DNA data (ATGX) to be end user report (patient report) that Doctors, researchers can use to consult to clients, other doctors, hospitals
- ✓ Be the main scientist to build the following products of Genesolutions which contribute apart to make them become a Billion startup in APAC: carrier4, oncosure, carrier1, carrier5, Whole Genome, rare diseases, NIPT research
- ✓ Presentation & sharing knowledge, creating attractive and challenging environment for the team
- ✓ Implement disease classification models to classify signals to 3 types of HBA diseases with Logistic regression, SVM, Random Forest. Fine-tune machine learning models
- ✓ Design parallel data Pipeline Architecture for running workflow Big Data – Human Genome WGS, Data analysis, data visualization, parallel programming with Nextflow tools. Create Automation system to run pipelines, expert in Docker, and Micro service architecture
- ✓ Experience with Linux servers, build GPU Application, Big Data (genome), cluster with HDFS – Spark, PySpark for BigQuery

- **Skills:** AWS sagerMarker, aws s3, lambda, IAM, mlops, ml pipelines. Docker, create machine learning architect app, linux

❖ **Project: Abnormal Variants Diagnose – NIPT**

- **Scope of work:**

- ✓ Main scientist to do NIPT Research to detect new variants in Vietnamese population of 1M

- people
- ✓ Build data pipelines to run 200TB fastq files (DNA data) to be meaningful variants
- ✓ Practice Big query using spark, build a clusters to run data pipelines (10 servers ram 1TB, 256 threads, a cluster GPU of 128GB)
- ✓ Work with NVIDIA, ILLUMINA team to build up Whole Genome Pipelines, using their software, hardware to accelerate GS's pipelines performance
- **Skills:** Data pipelines, Automation system, Linux, MacOS, docker, nextflow, cluster, big data, big query, research, statistic, write scientist papers

❖ **Achievement at GENESOLUTIONS:**

- ✓ Work with international team in NVIDIA, ILLUMINA to improve data pipelines
- ✓ Work directly with C-levels to implement the key products
- ✓ Manage timelines, tasks, complete before timeline, work with Senior, PHD scientist, doctors, high educated team

OUTSTANDING POINTS

- Proficiency in Python, R, Bash Script (Linux), C++, Git version control
- Advance data visualization with Plotly, **Streamlit**
- Advance skills: AWS Sage Marker, EC2, Azure, blob storage, AzureML, Restful API, Flask, Django Python, MySQL/Mongo DB, Deep learning, NLP, Pytorch/Tensorflow, Modeling, Data visualization
- The **best Employee Performance** in Gene Solutions 2020
- Manage and work with international teams well under high pressure
- Strong in Logic-Math problems, Statistics, Experimental Design, Hypothesis Testing, System Design
- Proficiency in English - **IELTS 6.5**, French – **DELF B1** equivalent

PUBLICATION

- Science Research: Wind turbine emulator based on DC motor DSP DSPACE 1103 research and design, paper [1]
- Android-Arduino Application for Photovoltaics Water Pump System paper [2]
- Speech Recognition using deep learning method, paper [3]
- Massively parallel sequencing uncovered mutational spectrum of three common metabolic diseases in Vietnamese pregnant women [4]

REFERENCES

- Ph.D. M. Cheng Lili, Postdoctoral Research Fellow, lilucheng@carro.sg, (+65)-90186790
- Ph.D. Giang Hoa, Postdoctoral Researcher, hoagiang@genesolutions.vn, (+84)-918933242