

# GEOFFREY WONG HIN

## Full-Stack Data Scientist/ ML Engineer

Personal Website: <https://geoffreywonghin-personal.vercel.app>

+1 825 9457174 | [geoffreywonghin@gmail.com](mailto:geoffreywonghin@gmail.com) | [www.linkedin.com/in/geoffreywonghin](http://www.linkedin.com/in/geoffreywonghin) | [github.com/RyanStark223232](https://github.com/RyanStark223232)

### RELEVANT SKILLS

---

- Proficient in Pytorch, Scikit-learn, MLFlow | Python, SQL | GCP, Azure | Apache Beam | Looker Studio, Tableau, PowerBI | Airflow, Gitlab-CI, Docker | FastAPI
- Experience with Keras, TensorFlow, TensorFlow Extended | R, C, Java, DAX | AWS | Hadoop, Spark, Databricks | Looker | Django | React JS | Flask | RAG, LoRA
- **Managed** the development of two data dashboards and a **Machine Learning client project**. Completed projects on time despite tight schedules and captivated clients with explorations of future directions
- Demonstrated **proactive** problem-solving skills by independently identifying inefficiencies in data pipelines and devising solutions with heuristic algorithms that improved **scalability and memory usage by 14x** and reduced the **processing time by 50%**
- Investigated the use of **LLM** in Personally Identifiable Information (PII) cleaning and web content understanding, and created systems that outperformed manual labeling services in speed and robustness

### RELEVANT WORKING EXPERIENCE

---

#### RESEARCH ASSISTANT (Part-time) & INDEPENDENT DEVELOPER

THE CHINESE UNIVERSITY OF HONG KONG & SHATIN HOSPITAL

08/2023 - 08/2024

- **Parkinson's Diagnosis Research:**
  - Designed collection procedures for patient data that satisfy researchers' demand for efficiency while maintaining a controlled experimental environment
  - Experimented with off-the-shelf AI products to eliminate **personal information**, reduce the **complexity** of the collected videos, and ensure prediction output did not **overfit** despite a small dataset
- **Natural Language Query Interface:**  
Built chatbots capable of querying databases, answering questions with accurate information, and generating graphs using an **agent-based Large Language Model (LLM)**

#### DATA SCIENTIST

HOTMOB LIMITED

07/2021 - 06/2023

- Leveraged **web-scraping** and transfer learning with **GPT-2** for comprehensive website understanding. Extended capability to sentiment analysis, interest classification, and keyword extraction, while increasing **capacity by more than 10 times with half of the cost**
- Impressed the Business Intelligence team of a global insurance company with client projects, including **revamping their data processing system** originally in SAS to PowerBI, training machine learning models to predict insurance fraud with **Azure**, and adding **explainable ML** to inference output, establishing long-term collaborative relationships to this day
- Fully automated data processing pipelines using **Airflow**. Removing the need for manual labor in monitoring data quality, updating ML models, and generating reports
- For more details, visit [github.com/RyanStark223232/work-portfolio-hotmob](https://github.com/RyanStark223232/work-portfolio-hotmob)

#### AI DEVELOPER INTERN

FLYING MILK TEA LIMITED

07/2020 - 09/2020

- Performed **Web Crawling** with **Selenium** to Collect Images and 3D Model Data

- Preprocessed 3D Mesh Data with **Programmable Blender**
- Built a **Voxel Generator** that Converts 2D Images into 3D Models

## EDUCATION

---

### MASTER of DATA SCIENCE & ANALYTICS (Health Data Science & Biostatistics)

UNIVERSITY OF CALGARY

09/2024 - PRESENT

### BACHELOR of SCIENCE in COMPUTER SCIENCE (Artificial Intelligence)

THE CHINESE UNIVERSITY OF HONG KONG

09/2017 - 07/2021

GPA: 3.85 / 4

#### Scholarship/Award:

- Gold Award for Outstanding Academic Performance
- Prof SC Loh Scholarship for CSE
- Li Po Chun Charit T Fund UGD
- Yu To Sang Memorial Scholarship 2019/20
- ELITE Stream Scholarship
- CodeVita 2018 Hong Kong 1st runner-up (Python & C )

#### Projects:

- Developed a Computer Vision system for Parkinson's Disease diagnosis that outperformed previous techniques **(Final Year Project)**
- Built an Ensemble Atrial Fibrillation classifier with accuracy comparable to the 2017 PhysioNet Competition winner **(Individual TensorFlow Project)**
- Evaluated the randomness of card shuffling techniques using Recurrent Neural Networks and a simulated card guessing game. Found room for improvement for existing algorithms **(Group TensorFlow/Programming Project)**
- Wrote a Review Paper comparing 10 different machine learning classifiers for pre-mRNA splicing prediction **(Individual Bioinformatic Project)**

## PROFESSIONAL DEVELOPMENT & CERTIFICATES

---

### Professional Machine Learning Engineer

GOOGLE CLOUD CERTIFIED

06/2023 - 06/2025

### Professional Project Management Certificate

GOOGLE

07/2023 - PRESENT

### Google's Large Language Model (LLM) Learning Path

CLOUD SKILL BOOST

06/2023 - PRESENT

### Dive into Deep Learning Using MXNet Workshop

AWS

08/2019 - PRESENT