

LOW WEI HONG

404.247.2157 • Atlanta, GA • lowweihong14@gmail.com • linkedin.com/in/lowweihong • github.com/lowweihong • Open to Relocation

Senior Data Scientist - NLP

Senior Data Scientist with over 6 years of expertise in production-grade machine learning and NLP, specializing in real-time data pipelines, feature engineering, and LLM development. Proficient in Python, PyTorch, and AWS, I excel in deploying scalable AI solutions and optimizing models for high accuracy. Currently pursuing an M.Sc. in Computer Science (Machine Learning) at Georgia Tech, I am eager to advance AI innovation as an Artificial Intelligence Engineer.

Key Skills

Machine Learning: Natural Language Processing (NLP), Deep Learning, Transformers, PyTorch, TensorFlow, Retrieval-augmented generation (RAG), Supervised Learning, Unsupervised Learning | Programming: Python, SQL, C++, Solidity | Cloud & DevOps: AWS, Docker, Airflow, Kafka | Big Data: Spark, Hive, Hadoop | Web Scraping: Scrapy, Playwright, Selenium | Data Visualization: Tableau, Plotly, Matplotlib | Others: Web3, Smart Contracts, Rapid Prototyping, Cross-functional Collaboration

WORK EXPERIENCE

Georgia Institute of Technology - Atlanta, GA, USA

03/2025 - 05/2025

EIC Lab Research Intern

- Developed LLM-based automation for Verilog code, streamlining issue resolution in repositories using SWE-bench-style techniques for enhanced code quality.
- Engineered Retrieval-Augmented Generation (RAG) workflows, optimizing Verilog task efficiency through advanced AI-driven solutions.

OKX - Singapore

11/2022 - 08/2024

Senior Data Scientist

- Engineered a KYC passport validation tool by deploying Optical Character Recognition (OCR) to extract data and sophisticated text-matching logic to verify consistency across both sides of the document, resulting in 99% matching accuracy.
- Engineered and deployed an internal generative AI chatbot using Faiss (Vector DB), automating responses for P2P merchants and reducing manual support workload by 35%.
- Owned the core address matching component for a user-wide risk monitoring system, using embedding matching method to achieve 95% accuracy in identifying potentially linked accounts. The model's output, a similarity score, served as a critical input for the overall risk factor calculation.
- Fine-tuned and deployed more than 13 translation models, cutting vendor translation costs by 35%.
- Optimized translation model inference speed by 1.5x using CTranslate for GPU-based deployment.
- Built and deployed a news intelligence distribution system, achieving more than 95% accuracy in matching news to internal teams.
- Designed and deployed a centralized web crawling system using Docker, enabling 40+ concurrent crawlers across multiple markets with zero downtime for 10+ users.
- Automated lead generation from YouTube crypto influencers, delivering leads to global BD teams.
- Scraped and structured data from copy-trade platforms, powering 10+ dashboards used to monitor trading performance of 57K+ KOLs and support business development targeting.
- Collected near real-time data from X.com (Twitter), enabling 100% recall of user complaints and triggering real-time alerts during surges to support timely customer service response.

Great Eastern Life – Singapore

06/2022 – 10/2022

Senior Data Scientist

- Developed a machine learning model for AML (Anti-Money Laundering) compliance, leveraging supervised and unsupervised techniques to achieve a 90% recall rate on fraudulent transactions.
- Discovered the main complaints of customers by clustering ~1mil of email by using BERTopic (topic modelling).

Hackwagon Academy - Singapore

06/2020 – 10/2020

Data Science 101 Instructor

- Developed and refined curriculum for a 7-week intensive data science course covering Python, statistics, and data visualisation, tailored specifically for beginners with no prior programming experience.
- Instructed and mentored over 100 adult learners across multiple cohorts, guiding them from foundational concepts to building complete data analysis projects.

- Explained complex statistical and programming topics in an accessible manner, fostering an engaging learning environment that led to high student satisfaction and successful course completion.

Shopee - Singapore

06/2019 - 06/2022

Senior NLP Data Scientist

- Trained spam detection models, achieving 98%+ precision across 6 markets in production.
- Developed a classification system for Indonesian marketplace listings, categorizing 3,000+ categories with 80%+ accuracy.
- Extracted product keywords via text mining, securing 40%+ acceptance by internal reviewers.
- Built SPU classification models for laptops and maintained over 97% accuracy across regions.
- Deployed and maintained 7 daily ML pipelines using Airflow for production monitoring and updates.

Shopee - Singapore

08/2018 - 05/2019

Tech Lead, Local Business Intelligence

- Trained machine learning models for product classification, achieving 97% accuracy and reducing human annotation by 60%.
- Single-handedly led the LPG initiative end-to-end—from web crawling to data pipelines and database management—improving system efficiency and reducing LPG claim rates by 50%.
- Automated sending report process from uploading file to google drive, sending email and updating google sheet.
- Mentored two interns, converting one to a full-time role through targeted skill development.

Shanghai Linksure Network Technology Co. Ltd – Singapore

12/2017 - 05/2018

Data Analyst Intern (Part-Time)

- Queried over 1 billion big data in Hive to analyse users' distribution within a time frame.
- Created a json parser to give insight about the housing market in a specific region in China using Python.
- Applied machine learning method to cluster factories into regions for top 10 cities in China.
- Built a multi-threaded Python web crawler, collecting millions of POI's (Point of Interest) and hundreds of thousands of AOI's (Area of Interest) efficiently.

Dentsu Aegis Network, Global Data Innovation Centre – Singapore

05/2017 - 07/2017

Data Science Intern

- Proposed and implemented a streamlined Python automation process to convert ontologies from XMind to OWL format, shortening the workflow from hours to minutes.
- Built a Python script to automate Google Trends calculations and comparisons for multiple companies, aiding competitive analysis.
- Designed marketing ontologies using XMind for visualization and Protégé for classes/relationships, boosting campaign knowledge representation.

PROJECTS

Permissionless IV Hackathon – Brooklyn, NY, USA

06/2025

Winner, Flow Blockchain Track

- Co-developed a Web3 platform to enable creators to license their voices via smart contracts and earn royalties from AI-generated audio.
- Tech Stack: Python (Flask), Docker, Nodeops, Flow Blockchain (Smart Contracts, USDC), IPFS.

Research & Publications

- H. Huang, W. Low, D. Zhang, et al. (2025). "EIDSeg: A Pixel-Level Semantic Segmentation Dataset for Post-Earthquake Damage Assessment from Social Media Images" *Submitted to the AAAI Conference on Artificial Intelligence (AAAI), Artificial Intelligence for Social Impact Track*.

EDUCATION

• M.Sc. in Computer Science (Machine Learning) Georgia Institute of Technology, USA

12/2025 (Expected)

• B.Sc. (Hons) in Mathematics and Economics (Double Major) Nanyang Technological University, Singapore

07/2018

CERTIFICATE

• NLP with Deep Learning, Stanford Online

12/2020

• Executive Program in Algorithmic Trading, QuantInsti

08/2018

LANGUAGE

Languages: Mandarin (Native), English (C2), Malay (Intermediate)