

# Fei Huang

Amsterdam, Netherlands | +31 645 038 614 | huangfo6@gmail.com  
[linkedin.com/in/huangfo6](https://linkedin.com/in/huangfo6) | [github.com/huangfo6](https://github.com/huangfo6) | [feithink.substack.com](https://feithink.substack.com)

---

Data Engineer with expertise in building scalable data platforms and ETL pipelines. Skilled in Python, Databricks/PySpark, and cloud technologies. Hands-on experience designing data infrastructure and optimizing pipeline performance. M.Sc. in AI from VU Amsterdam. Seeking to contribute to Picnic.

## EDUCATION

---

### Vrije Universiteit Amsterdam, The Netherlands

M.Sc. in Artificial Intelligence (GPA: 8.2/10)

Amsterdam

Sep 2023 – Aug 2025

*Thesis: Uncertainty Quantification in Deep Reinforcement Learning under Noisy Environments*

### Tsinghua University, China

B.Eng. in Industrial Engineering

Beijing

Sep 2006 – Jul 2010

**Certification:** Databricks Certified Data Engineer Professional (Feb 2026)

## PROFESSIONAL EXPERIENCE

---

### GLP Technology

Founding Data Engineer & Risk Lead

Shanghai, China

Jul 2017 – Aug 2019

- Joined as founding data team member at fintech startup; built credit scoring engine from scratch including data pipelines, feature engineering, and model deployment using logistic regression and scorecard methodology.
- Led PySpark workflow implementation to process large-scale datasets, reducing processing time and enabling daily risk reporting.
- Engineered automated data pipelines using Python and SQL; integrated payment gateway APIs and coordinated collection operations.

### Baiquan Investment

Quantitative Developer

Beijing, China

Jul 2015 – Jun 2017

- Developed the underlying event-driven backtesting infrastructure that supported strategy simulation and performance reporting.
- Engineered high-performance factor computation engine using vectorized Pandas/NumPy operations; reduced calculation time for technical indicators across 3000+ stocks.
- Built automated ETL pipeline to ingest daily market data from vendors with data quality validation supporting downstream trading operations.

### Ele.me (acquired by Alibaba)

Data Analyst

Shanghai, China

Sep 2013 – Jul 2015

- Built user profiling and classification model for marketing campaigns that doubled recall rate of churned customer targeting through A/B testing.
- Queried and processed large-scale user data from Hadoop cluster using Hive SQL; generated reports and dashboards to support business decision-making.

*Career Note: 2019–2023 included independent investing, language learning (English, German), and graduate preparation.*

## PROJECTS

---

### Financial Data Lakehouse (Databricks/Spark)

*Oct 2025 – Present*

- Architected resilient lakehouse using Auto Loader and Structured Streaming; pipelines handle schema evolution automatically for continuous crypto market data ingestion.
- Engineered data quality framework via fault injection testing; validated quarantine pattern to isolate malformed records without disrupting production.

### M.Sc. Thesis: Uncertainty Quantification in Deep RL

*Feb 2025 – Aug 2025*

- Built evaluation framework for uncertainty quantification in deep RL: implemented calibration metrics (ACE, CRPS, coverage) and evaluated model reliability under distribution shifts.
- Developed calibration methods including temperature scaling and bias correction to improve model reliability under uncertainty.

## TECHNICAL SKILLS

---

**Programming:** Python (pandas, NumPy, PyTorch, scikit-learn), SQL, Bash

**Data Infrastructure:** Databricks, PySpark, Delta Lake, Apache Airflow, Kafka, Hive

**Cloud & DevOps:** AWS, Azure, GCP, Docker, CI/CD, Terraform, Git

**ML & Statistics:** XGBoost, LightGBM, Time Series Analysis, A/B Testing, Statistical Modeling

## INTERESTS

---

**Interests:** Philosophy (Kant, existentialism), Dostoevsky, powerlifting, analytical writing

**Selected essays:** [huangfo6.github.io/FeiThink/en/](https://huangfo6.github.io/FeiThink/en/)