

HAOYANG HAN

haoyanghan1996@gmail.com | +86 139-5320-7399 | Huangpu District, Shanghai

PROFESSIONAL SUMMARY

- LLM-focused Senior Data Scientist with 5.5 years of experience designing and deploying machine learning and GenAI solutions from PoC to production across finance, NLP, and enterprise use cases.
- Deep expertise in RAG pipelines, prompt engineering, LLM evaluation, and LLM-as-a-judge strategies.
- Proven track record in leading 4–6 person cross-functional teams and presenting to executive leadership. Adept at international collaboration with multi-time zone stakeholders.

RELEVANT EXPERIENCE

Citigroup, Innovation Lab, Shanghai, China

Senior Data Scientist, Assistant Vice President

Aug. 2021 – Present

GenAI & NLP Leadership

- The **most senior** LLM/GenAI data scientist in Citi China; led multi-disciplinary teams (4–6 members) including junior data scientists, PMs, and engineers locally and globally.
- Built production-ready LLM solutions from scratch including:
 - Memo generation system (Orion) using customized RAG with Llama-Index. 65% time savings on generation, **\$40 Million** projected annual savings.
 - LLM evaluation & prompt enhancement platforms (Prompt Hub) using meta prompts and LLM-as-a-judge methodology. → Convert prompts cross models, saved **70%** prompt composing time, **40%** prompt quality improvement on average, 300 users.
 - Enterprise-scale search engine (CitiFindr) → 20K+ users, 4M+ interactions in 10 months, **\$9M projected annual savings**.
 - Search and memo generation pipeline for frontend wikipedia (BankingWiki), which contains image based pdfs(vision embedding), complicated tables, and super large volume document for retrieval. In production for entire HongKong front desk.
 - Question answer generation & evaluation system using RAG (Data Curation) → Fine-tune and risk management data generator, ready for production, used by more than **9 use case teams** to generate 5,000+ QA pairs.
 - News-to-signal classifier (PSDS): In prod for 4 years, detected Russia-Ukraine war 3 minutes ahead global media.

Platform vs. Use Case Roles

- Participated on entire Citi Data Science lifecycle in both platform and use cases.
 - Platform side: On Citi's LLM platform "Stellar" — evaluated and determined model onboarding, established usage guidelines across departments; design Stellar to C roadmap and delivered Data Curation and Prompt Hub.
 - Use case side: Lifecycle data science project involvement. Deployed and monitored task-specific models for banking, audit, and operations workflows.

Global Collaboration & Leadership

- Led technical collaboration across NYC, London, Tel Aviv, Dublin, aligning product design and data regulations.
- Demoed solutions to executive leadership, including Citi CEO Jane Fraser's direct reports.
- Conducted internal educational sessions on GenAI domain like Google Learn About, Deep Research, and LLM-as-a-Judge.

Evaluation, Risk, and Governance

- Owned model evaluation and risk governance: designed metrics, gathered user feedback, managed compliance communications on 8 projects, which is both experimental design and output quality assurance.
- Collaborated with regulatory, audit, and business units to ensure GenAI model alignment with financial compliance.

Proof of Concept Solution Architecture & Deployment

- Designed full-stack data/LLM pipelines for:
 - Agentic workflow toolsets for daily enterprise use: Jira Ticket Generator, document to PPT converter, Email Polisher. Was piloted and used by 200 users.
 - Knowledge Graph based information retrieval system and a fine-tuned seq-2-seq transformer for Sparql(SQL-like) query generation and semantic search.

Earlier Responsibilities

- Managed data science initiatives across internal audit, market ops, info security, and investment banking teams before GenAI specialization.
- Developed end-to-end solutions, gaining full lifecycle experience in financial data science.

Internet Brands, Inc.(sub affiliate of K.K.R.), Los Angeles, U.S.

Associate Data Scientist

Mar. 2019 – Jun. 2021

- 1. Developed an NLP pipeline (SILK) for auto-article suggestion by topics, keywords, and keyword values to optimize the website ranking in search engine (SEO).
- Generated keywords from article using PositionRank, encoding the generated keywords and the target keywords by pre-trained transformers (BERT), and calculate the cosine similarity for final recommendation.
- Deployed SILK model to 7 websites across Internet Brands, with more than 400k articles involved. Coordinated with multiple develop team for regular time-based deployment and refresh using crontab.
- 2. Developed churn model for Martindale Hubbell website using salesforce data and random forest algorithm. Making prediction for 2600 customers based on past 1-year data and did hyper-parameter tuning to achieve 0.78 AUC score.
- 3. Generated summaries of legal websites (Ngage, lawyers.com)' chat transcripts by both constructing an encoder-decoder system with GloVe embeddings, Bi-directional GRU and using pre-trained models like BERT and XLNet.
- Deployed model to production with trained and pre-trained model using virtual environment and flask for real time text summarization and area of purpose (AOP) prediction.

Alpha2Fund (Hedge Fund), Beijing, China

Intern, Machine Learning Quantitative Researcher

Jun. - Sep. 2018

- Designed new Alpha factor combination pipeline by implementing ensemble machine learning algorithms, encompassing LR, RF, MLP, kNN, XGBoost, for decision making in stock market, utilizing hundreds of alpha factors in commodity future market.
- Improved information correlation from 1.16 with the linear combination baseline to 1.57 by forecasting long/short position of specific stocks by training and GRU (for multiple stocks) on stock alpha factors.
- Accelerated minute level stock prediction by 20% by replacing XGBoost based model with a lightGBM based model for factors combination.

EDUCATION

Northwestern University, Evanston, Illinois

M.S. in Project Management | GPA: 3.74 out of 4.00

Sep. 2017 - Dec. 2018

- Courses: Machine Learning: Foundations, Algorithms, and Applications; Deep Learning Foundations from Scratch; Introduction to Database Systems; Deep Reinforcement Learning; Time Series Analysis; Deep Learning Individual Project.

Ocean University of China

Bachelor of Arts (BA), Marine Resource and Environment

Sep. 2013 - Jun. 2017

SKILLS

- Languages and packages: Python, SQL, Bash, Javascript yTorch, HuggingFace Transformers, LlamaIndex, Docker, Git, Jupyter
- LLM Expertise: MetaPrompt, RAG, Agentic Workflows, LLM-as-a-Judge