JONG INN PARK

+1 763-340-4178 | park2838@umn.edu | Minneapolis, MN

SUMMARY

Data-driven **Data Scientist/LLM Engineer** with research cited **215+** times, known for turning unstructured data into measurable business outcomes. Built **automated workflows** that saved **$200K annually** and unlocked **$20M+** potential **cost avoidance** by scaling **compliance monitoring** with computational linguistics; cut **25–40** staff hours/quarter by analyzing **300–500** earnings calls with topic modeling & sentiment; and standardized leadership coaching for **100+** U.S. Naval Academy trainees via an AI chatbot that converts conversations into trackable KPIs. Proven ability to create scalable AI solutions adopted across **academia** and **industry**, from peer-reviewed AI research to **financial services** and **education**.

SKILLS

- **Programming Languages:** Python, R, SQL, TypeScript/JavaScript, Java
- **Web:** FastAPI, Flask, Vue, React, D3, HTML/CSS, REST, RabbitMQ/Redis/Celery, Node.js, Postman
- **ML/LLM:** PyTorch, HuggingFace, LangChain, LangGraph, Scikit-learn, Pandas, Numpy
- **Data:** ER Modeling, PostgreSQL, MongoDB, MySQL, Chroma, ETL, Scrapy, Matplotlib, Seaborn
- **DevOps:** Docker, Nginx, Ngrok, Git, GitHub Actions, Alembic, Pytest, Flower, CI/CD
- **Cloud & Analytics:** AWS, Azure, Streamlit, Tableau, Excel, Confluence

EXPERIENCE

Full Stack Web Developer Jul 2024 – Present

TNTLAB, Minneapolis, MN, USA

- Deployed a leadership coaching AI chatbot used by 100+ U.S. Naval Academy trainees, converting subjective, unstructured conversations into standardized, trackable outcomes; validated a model positioned to scale into school-based advising and mental healthcare.
- Automated reminders, tracking, and reporting, reducing reliance on manual oversight and unlocking large-scale time and cost savings through continuous program evaluation.
- Built scalable infrastructure (databases, logging pipelines) that enabled new cohorts and programs to be onboarded with minimal engineering effort, reducing deployment costs and ensuring sustainability.

Research Engineer Apr 2023 – May 2025

Minnesota NLP Lab, Minneapolis, MN, USA

- Pioneered data pipelines and analysis workflows for ACL (Association for Computational Linguistic) 2024 study on cognitive biases in LLMs; gained 160+ citations in its first year, demonstrating broad adoption and impact.
- Designed an agentic video-generation workflow with LangChain, MoviePy, and LLaVA to produce 45 short-form science videos; evaluation showed +25% higher clarity and alignment than baseline, expanding reach of science communication.
- Engineered a framework (SelectLLM) that automated high-value instruction selection, enabling efficient LLM fine-tuning with smaller datasets and lowering time and cost barriers for industry use; research already cited 17 times in academic and applied contexts.
- Partnered with Linguistics & Psychology departments to build evaluation pipelines for narrative understanding in LLMs; project recognized at ACL 2025 and reused across multiple labs, highlighting cross-disciplinary scalability.
- Built a lightweight Python platform (talkative-llm) to unify inference across models (e.g., LLaMA, GPT-4); adopted in two graduate-level NLP courses to help 40+ students learn rapid prototyping with LLMs.
- Developed Chrome extensions, APIs, and dashboards to capture 1,000+ real-world search logs; delivered insights with Journalism faculty that cut ad analysis turnaround by ~1 week and authored AEJMC-accepted research.
- Containerized and deployed full-stack web apps with Docker, enabling researchers and law students to run structured LLM experiments and simulate 20+ legal consultations (~100 minutes each) in production settings.

Data Science Intern Sep 2023 – Apr 2024

Piper Sandler, Minneapolis, MN, USA

- Automated analysis of 300–500 quarterly earnings-call transcripts, applying topic modeling and sentiment analysis to highlight key themes and tone; reduced manual review time from ~5 minutes per call to near-instant scanning, saving analysts 25–40 staff hours each quarter.
- Delivered a Streamlit-based prototype app that let Equity Research teams quickly search, visualize, and export insights from earnings calls, improving turnaround time for investment decision support.
- Built Python (Scrapy) pipelines on the SEC EDGAR API to automatically ingest 10-K and 10-Q filings, standardizing fields and sections into a normalized schema; reduced repetitive manual review across filings, enabling Equity Capital Markets teams to identify key disclosures faster and focus on higher-value analysis.

Data Analyst Jan 2019 – Jul 2022

Samsung Card, Seoul, South Korea

- Saved $200K annually and unlocked $20M+ potential cost avoidance by developing and deploying NLP-based classification models (BERT, Naive Bayes, SVM) on 100k+ monthly debt-collection call transcripts, scaling compliance monitoring from 1% to 100%.
- Eliminated manual effort equal to 300 FTEs and reduced compliance breaches by integrating speech-to-text pipelines and active-learning model training, shifting agent behavior under full monitoring awareness.
- Cut reporting time by 80% by building SQL + ETL pipelines to aggregate 3,000+ daily calls and designing executive-ready dashboards (Tableau, internal BI tools) for anomaly detection and staffing optimization.
- Accelerated campaign planning by ~2 weeks through Python-based web scraping and NLP (NER, sentiment, keyword extraction), delivering 12 monthly trend reports that directly guided merchandising managers in product decisions.
- Reduced competitive analysis cycle from weeks to days by standardizing 300+ competitor credit-card offerings using Python, SQL, and schema normalization techniques, linking external product data with internal transaction databases.
- Processed 1,000+ trending products weekly by engineering batch data pipelines (Python, Cron jobs) across 10+ social and e-commerce platforms (e.g., Instagram, Coupang), transforming 500+ posts and 1,000+ replies into structured datasets.

EDUCATION

Master of Science in Data Science

University of Minnesota, MN, USA Sep 2022 - Jun 2024

Bachelor of Economics in Statistics & Bachelor of Science in Systems Management Engineering

Sungkyunkwan University, Seoul, South Korea Mar 2013 - Feb 2019

PUBLICATIONS

- **How LLMs Comprehend Temporal Meaning in Narratives: A Case Study in Cognitive Evaluation of LLMs**
  ACL 2025 Main (May 2025) Paper
  Authors: Karin De Langis, **Jong Inn Park**, Andreas Schramm, Bin Hu, Khanh Chi Le, Dongyeop Kang
- **Stealing Creator's Workflow: A Creator-Inspired Agentic Framework with Iterative Feedback Loop for Improved Scientific Short-form Generation**
  Preprint (Apr 2025) Paper | Project Page
  Authors: **Jong Inn Park**, Maanas Taneja, Qianwen Wang, Dongyeop Kang
- **A Framework for Robust Cognitive Evaluation of LLMs**
  Preprint (Apr 2025) Paper
  Authors: Karin de Langis, **Jong Inn Park**, Bin Hu, Khanh Chi Le, Andreas Schramm, Michael C Mensink, Andrew Elfenbein, Dongyeop Kang
- **Consumer Engagement with AI-Powered Search Engines and Implications for the Future of Search**

Advertising
AEJMC 2024 Advertising Division (Aug 2024) Paper
Authors: Gabriel Garlough-Shah, **Jong Inn Park**, Shirley Anugrah Hayati, Dongyeop Kang, Jisu Huh

- **Benchmarking Cognitive Biases in Large Language Models as Evaluators**
  Findings of ACL 2024 (Aug 2024) Paper | Github | Project Page
  Authors: Ryan Koo, Minhwa Lee, Vipul Raheja, **Jong Inn Park**, Zae Myung Kim, Dongyeop Kang
- **SelectLLM: Can LLMs Select Important Instructions to Annotate?**
  Preprint (Jan 2024) Paper | Github
  Authors: Ritik Sachin Parkar*, Jaehyung Kim*, **Jong Inn Park**, Dongyeop Kang

LINK

- **LinkedIn**: https://www.linkedin.com/in/jong-inn-park/
- **My Page**: https://jong-inn.github.io/
- **GitHub**: https://github.com/jong-inn/
- **Google Scholar**: https://scholar.google.com/citations?user=3Hw6n5AAAAAJ&hl=en