LEO (JIAN-PENG) LI

+1 (650) 575-3574 | jianpengli1205@gmail.com | Personal Website | LinkedIn | Github

EDUCATION

Cornell University   New York, NY
Master of Engineering in Computer Science   Expected May 2026
- Cumulative GPA: 4.0/4.0 (Merit Scholarship)

University of California, San Diego   La Jolla, CA
Bachelor of Science in Mathematics-Computer Science; Economics; Minors in Data Science   Sep 2021 - Dec 2024
- Cumulative GPA: 3.8/4.0 (Errett Bishop Scholarship)
- Relevant Courseworks: Data Structures, Algorithms, Data Mining, Data Management, Operating Systems, Machine Learning, Deep Learning, Computer Vision, Recommender System, Optimization

WORK EXPERIENCE

Cathay Financial Holdings   Taipei, Taiwan
Machine Learning Engineer Intern (Digital, Data & Technology Center)   Oct 2024 – Present
- Utilized GCP BigQuery to preprocess ICD datasets with over 50 million entries, reducing training time by 70% and applying scaling weights to address class imbalance, increasing the F1 score by 0.2
- Implemented a BioBERT-based multi-label classification model with PyTorch for text analysis, expanding the label set from 30 to 50 and conducting error analysis on threshold sensitivity, which increased recall by 0.28

Anivance AI   Hsinchu, Taiwan
Research and Development Engineer Intern   Oct 2024 – Present
- Designed and developed a biomedical application integrating domain-specific AI agents with AutoGen, enabling internal agent communication for generating research insights on organ chip technology
- Implemented a multimodal retrieval system enabling both text and image retrieval, and conducted experiments on common embedding models, including OpenAI embeddings, BioBERT, and PubMedBERT, to identify the most suitable structure
- Built a retrieval pipeline, integrating GraphRAG and real-time Google Scholar scraping, enhancing retrieval accuracy by 25%

Shang Data Lab, UCSD   La Jolla, CA
Undergraduate Research Assistant   Feb 2024 – Present
- Designed a data generation pipeline to create semantically rich, long-context conversation data that simulates real-world scenarios for novel Memory-Bank model research, improving retrieval accuracy by 0.3 compared to similarity-based retrieval
- Evaluated advanced Large Language Models' (LLMs) reasoning capabilities in long-context settings, with token sizes ranging from 32k to 128k

CSE / Halıcıoğlu Data Science Institute (UCSD)   La Jolla, CA
Computer Science and Data Science Teaching Assistant   Sep 2023 – Dec 2024
- Assisted 1,000+ students in Computer Science topics, including Object-Oriented Programming (OOP), Data Structures, and Algorithm Design
- Achieved 92% recommendation rate and student satisfaction score by delivering content in a clear, understandable manner

Shang Chin Education Co., Ltd   Taichung, Taiwan
Software Development Intern   Jun 2023 – Sep 2023
- Developed an AI-powered Chinese tutoring application by integrating the ChatGPT API with React.js and Node.js, enabling dynamic and interactive language learning
- Built a Chinese handwriting recognition application using JavaScript, CSS, and an SDK, improving teaching efficiency and adopted by 20+ online tutors

LEADERSHIP ACTIVITIES

Triton Quantitative Trading   La Jolla, CA
Web Development Team Lead / Founding Board Member   Jun 2023 – Oct 2024
- Directed the design and development of the club's fullstack website using React.js, MongoDB, and TypeScript
- Secured sponsorships from Jane Street Capital and the CSE Department

SKILLS

- Programming Languages / Databases: Python, Java, JavaScript, C++, SQL, MongoDB, MySQL
- Frameworks / Tools: PyTorch, TensorFlow, Pandas, NumPy, LangChain, AWS, GCP, Spark, Linux, React, Node.js, Git