Xuan Jiang

99 JOHN ST, New York, NY, 10038

+1 571-426-9968 xuanj@mit.edu xuan-jiang github.com/Xuan-1998 https://xuan-1998.github.io/ Scholar

Xuan is a software engineer at Google holding a Ph.D. degree from the **University of California, Berkeley**, specializing in Computer Science with a keen interest in **AI, High-Performance Parallel Computing (HPC), Large Language Model (LLM) post-training**.

Academic Education

University of California, Berkeley **June 2021 – Dec 2024**

Doctor of Philosophy, Specialized in Machine Learning and HPC *Berkeley, CA*

Courses: Applications of Parallel Computers, Introduction to ML, Back-End Web Architecture, Software Dev, Advanced Matrix Computations, Distributed Systems, Deep Reinforcement Learning and Decision Making

University of California, Berkeley **July 2020 – May 2021**

Master of Science, Engineering *Berkeley, CA*

Courses: Data Structures and Algorithms, Introduction to Database Systems, Introduction to AI, Operating Sys

Tongji University **Sep. 2016 – May 2020**

Bachelor of Science | Object-Oriented Design, Compilers, Internet Protocols *Shanghai, China*

Research Experience

Massachusetts Institute of Technology, Cambridge, MA **May 2023 – Now**

Research Affiliate in JTL - Applied RL approaches to design and evaluate optimization strategies in complex system environments, leveraging HPC based simulators to validate scalability and efficiency. Supervisor: Jinhua Zhao and Haris Koutsopoulos

Conducted research on LLMs and LLMSys, focusing on improving scalability and performance of training and inference, as well as advancing interpretability through specialized system design.

Pioneered research in developing LLM-based agents for anomaly detection in large-scale cloud systems, including integration of neuro-symbolic mechanisms to jointly process structured metrics and textual logs, symbolic verification for hypothesis validation, and structured reporting for interpretability. Built benchmarking datasets with fine-grained anomaly annotations and evaluated system performance in terms of accuracy, false positive reduction, and scalability.

Lawrence Berkeley National Laboratory, Berkeley, CA **June 2021 – May 2023**

Graduate Student Researcher in the EA - Energy Analysis Env Impacts Division. Spearheaded the development of predictive models for multi-modal systems, emphasizing the integration of machine learning techniques to optimize logistics and supply chain scenarios. Supervisor: Thomas Wenzel

University of California, Berkeley, CA **August 2020 – December 2024**

Research Scholar at Specialized Computing Ecosystems (SLICE) and Aviation Innovation Research (AIR) Lab. Develop GPU-based parallel multi-modal micro traffic operation simulator. Supervisor: Prof. Raja Sengupta. Co-supervisor Prof. Joan Walker, Prof. James Demmel, Prof. Mark Hansen, Prof. Daniel Rodriguez, Prof. Alexandre M. Bayen

University of Washinton, Seattle, WA **August 2020 – May 2023**

Research Assistant in Smart Transportation Application & Research Lab (STAR Lab). Contributed to pioneering research on city-wide network congestion prediction by integrating traffic science with representation learning, developing the Traffic-informed Transformer (TinT). This work involved overcoming challenges related to diverse sensor modalities, congestion fluctuation modeling, and network structure generalization. Played a key role in enhancing the accuracy of traffic forecasting through innovative approaches such as anisotropic graph aggregation and traffic-informed tokenization. Supervisor: Prof. Yinhai Wang

National Academies of Sciences, Engineering, and Medicine, Washington DC **April 2022 – Oct 2024**

Young Member in transportation research board standing committee on aviation safety, security and emergency

management (AV090) Review aviation-related papers, triennial strategic plan, and airport cooperative research program (ACRP) problem statements. Chair: Gaël Le Bris

Main Member in the Transportation Research Board Standing Committee on Marine Environment (AW030). Engaged in high-performance computing (HPC) and deep learning to address environmental challenges and promote sustainable practices in marine settings. Chair: Richard Billings

National Academies of Sciences, Engineering, and Medicine, Washington DC **April 2022 – Oct 2024**

Young Member in transportation research board standing committee on aviation safety, security and emergency management (AV090) Review aviation-related papers, triennial strategic plan, and airport cooperative research program (ACRP) problem statements. Chair: Gaël Le Bris

American Society of Civil Engineers (ASCE) **July 2023 – Oct 2024**

Committee Member in the Transportation Development Institute's Artificial Intelligence in Transportation Committee. Engage in activities related to AI applications in transportation systems, spanning across various transportation modes and aspects. Participate in conferences, workshops, and contribute to publications to foster innovation in AI applications within transportation. Chair: Lili Du, Ph.D., Aff.M.ASCE

Committee Member in the Active Transportation Committee. Participate in discussions, initiatives, and activities aimed at improving and promoting active transportation modes such as walking, biking, and other forms of human-powered transportation. Contribute to the development of best practices, guidelines, and policies that enhance the safety, accessibility, and convenience of active transportation. Chair: Cong Chen

Working Experience

Google **May 2024 – Present**

Software Engineer *New York, New York*

- Received the Software Engineering Evolution and Transformation (Sweety) Award for AI Pioneers, recognizing impactful contributions to advancing AI system design and large-scale applications.
- Selected as an Annual Review Outstanding Impact recipient for leading organization-wide AI initiatives with measurable engineering and productivity outcomes.
- Designed AI-assisted large-scale code change (LSC) workflows and AI-powered integration testing pipelines, significantly improving developer productivity (Accelerated one of the LSC efforts by 30+%) and system reliability at scale.
- Developed multi-agent frameworks for agentic code generation that integrate reasoning and execution, enabling robust automation for complex software engineering tasks.
- Initiated and launched the Google Cloud Assist (an LLM agent) alert-driven investigation workflow, enhancing incident diagnosis efficiency and operational responsiveness for troubleshooting.
- Used Google Cloud Services, e.g. GAE and Datastore, to design a GCP based infrastructure for Google internal probing system. It provided advanced I/O capabilities for verifying a system's ability to interact with various communication mechanisms, including Slack and Webhook. And now it's deployed in production for on-call team usage, which eliminates the pain of manual switching to a new workspace every month.
- Created OSS Notification Solution for Cloud Alerting Notification Forwarding solving our big customers' (e.g. Yahoo) pain points

California Consortium for Public Health Informatics & Technology **May 2023 – May 2024**

Artificial Intelligence in Public Health Informatics Software Developer *Berkeley, CA*

- Instructed over 250+ students in applying AI methodologies to Public Health Informatics and GIS Mapping & Analytics, emphasizing on the use of machine learning algorithms for spatial data analysis and public health decision-making.
- Acquired proficiency in leveraging LLM techniques within ArcGIS applications for Public Health, creating predictive models and analytics dashboards to forecast public health trends and challenges.
- Mastered the application of AI in enhancing data interoperability through RESTful API calls using HL7/FHIR resources on a Postman environment, automating data analysis and insights generation.

Lawrence Berkeley National Laboratory **June 2021 – June 2023**

Software Developer *LBNL-UCB-STI/beam/Xuan/ActivitySim-micromobility* *Berkeley, CA*

- Deploy agent based traffic simulation on AWS EC2 instances to enhance cloud computation and multi-task implement but also on NERSC(National Energy Research Scientific Computing Center) to utilize the Cray EX system with AMD CPUs and NVIDIA A100 GPUs Berkeley Lab owns

- Inherit the Akka FSM trait which provides a domain-specific language for programming agent actions as a finite state machine to achieve the goal of doing **traffic simulation based on 2,466,019** households travel plans from bay area

Shanghai HeroTech Education Technology Co., Ltd. **July 2020 – Sep 2021**

Co-founder *Shanghai, China*

- Dedicated to promoting public understanding and education in the fields of artificial intelligence, deep learning, and computer vision
- Recognized as a Microsoft Most Valuable Professional (MVP) in Artificial Intelligence and a Huawei Cloud Developer Expert (HCDE) in A.I.
- Member of the Popularization and Education Committee of the China Graphics Society
- Managed and grew a successful Bilibili channel, accumulating 210,000 subscribers and over 10 million cumulative video views
- Produced notable works, including in-depth discussions on AI research papers, comprehensive lectures on Stanford CS231N Computer Vision and CS224W Graph Neural Networks (in Chinese), and practical tutorials
- Engaged diverse audiences, including open-source developers, university students, postgraduate candidates, researchers, IT professionals, programmers, and individuals interested in machine learning, computer vision, AI art, and related fields
- Collaborated with renowned companies such as Huawei Cloud, Baidu, Alibaba, Amazon, SenseTime, and digital product companies on various projects and initiatives
- Established partnerships with publishers, open-source organizations, and participated in international events and exhibitions

China Ocean Shipping Company **July 2020 – Oct. 2020**

Data Engineer Intern *Shanghai, China*

- Developed and integrated an enterprise-level database, encompassing over **100,000+ shipping records**, showcasing proficiency in managing vast logistics datasets.
- Implemented the backend Restful API to provide up-to-date access to Google Earth and Kylin data.
- This API provided an intelligence utility that equipped **2000+ analysts** in the company.

Eastrong International Logistics Co., Ltd **Jan. 2018 – May. 2018**

Data Scientist Intern *Shanghai, China*

- Conducted an inventory forecast by building a linear regression model in Python based on historical data, which increases the **accuracy** by **17%** and further analyze the causality relationships between different factors.
- Validated the prediction on dataset, and developed a dynamic visualization dashboard with JavaScript and Tableau.

Journal Publications

Jiang, X., Sengupta, R., Demmel, J., and Williams, S. (**2024**). Large scale multi-GPU based parallel traffic simulation for accelerated traffic assignment and propagation. Accpted by *Transportation Research Part C: Emerging Technologies 169, 104873*.

Jiang, X., Cao, S., Mo, B., Cao, J., Yang, H., Tang, Y., Hansen, M., Zhao, J., and Sengupta, R. (**2024**). Simulation-based optimization for vertiport location selection: A surrogate model with the machine learning method. *Transportation Research Record: Journal of the Transportation Research Board*. Accepted Feb 2024.

Jiang, X., Tang, Y., Tang, Z., Cao, J., Bulusu, V., Poliziani, C., Sengupta, R. (**2024**). Simulating the Integration of Urban Air Mobility into Existing Transportation Systems: A Survey. Journal of Air Transportation 32 (3), 97-107

Tang, Y., Qu, A., **Jiang, X.**, Mo, B., Cao, S., Rodriguez, J., Zhao, J., Wu, C. (**2024**). Domain-Randomized Curriculum for Robust Reinforcement Learning in Bus Operations. *Smart Cities 7 (6), 3658-3677*.

Wenzel, T., **Jiang, X.**, Needell, Z., and Poliziani, C. "Simulating Docked Bikeshare and Public Transit in the San Francisco Bay Area." In *Energy Technologies Area, Lawrence Berkeley National Laboratory*, Berkeley, CA, USA, Nov **2023**. [Technical Report]

Yang, H., Zheng, W., Cai, J., Wang, P., **Jiang, X.**, Du, S., Wang, Y., and Wang, Z. (**2023**). Integrating the traffic science with representation learning for city-wide network congestion prediction. Information Fusion, 99, 101837. DOI: 10.1016/j.inffus.2023.101837

Bauranov, A., Parks, S., **Jiang, X.**, Rakas, J., & González, M. C. (**2021**). Quantifying the Resilience of the US Domestic Aviation Network During the COVID-19 Pandemic. Frontiers in Built Environment, 7, 642295. DOI: 10.3389/fbuil.2021.642295

Chai, C., Lu, J., **Jiang, X**., Shi, X., & Zeng, Z. (**2021**). An automated machine learning (automl) method for driving distraction detection based on lane-keeping performance. arXiv preprint arXiv:2103.08311. Submitted to IEEE Transactions on Pattern Analysis and Machine Intelligence, Aug 2022. Under review.

Conference Proceedings

Jiang, X., Zhuang, D., Cao, S., Cao, J., Tang, Y., Li, J., Bulusu, V., Sengupta, R., and Zhao, J. "Performance Benchmarking and Scalability of LPSim: A Multi-GPU Traffic Simulation Approach." In *the 2023 IEEE International Conference on Digital Twins and Parallel Intelligence*, Orlando, FL, USA[Poster Session]

Jiang, X., Yang, H., and Wang, Y. "Making Sense of Electrical Vehicle Discussions Using Sentiment Analysis on Closely Related News and User Comments." ***2022*** *International Conference on Transportation and Development (ICTD)*, American Society of Civil Engineers (ASCE). [Conference paper]

He, H., Li, J.B.,**Jiang, X.** and Miller, H. "Sparse Matrix in Large Language Model Fine-tuning." *ICLR 2024 Conference*, Singapore [Accepted Paper]

Jiang, X., Peng, X., Bulusu, V., Poliziani, C., Chatterji, G., & Sengupta, R. (2022, September). A Metrics-based Method for Evaluating Corridors for Urban Air Mobility Operations. In **2022** IEEE International Smart Cities Conference (ISC2) (pp. 1-7). IEEE. [Conference paper] DOI: 10.1109/ISC255366.2022.9922442

Tang, Y., Cui, K., Park, J. H., Zhao, Y., **Jiang, X.**, He, H., Yu, J. B., Koutsopoulos, H., and Zhao, J. "RAST-MoE-RL: A Regime-Aware Spatio-Temporal MoE Framework for Deep Reinforcement Learning in Ride-Hailing." *ICLR 2026 Conference* [Under Review]

He, H., Ding, X., **Jiang, X.**, Cheng, A., Zhao, Y., Li, J. B., and Miller, H. "MoECondenser: Finetuning MoE LLMs with Condenser Experts." *ICLR 2026 Conference* [Under Review]

Zou, X., **Jiang, X.**, Huang, R., He, H., Kapoor, P., Wu, H., Wang, Y., Sha, J., Shi, X., Huang, Z., and Zhao, J. "Towards Generalizable Context-aware Anomaly Detection: A Large-scale Benchmark in Cloud Environments." *ICLR 2026 Conference* [Under Review]

Zhao, Y., Qu, A., **Jiang, X.**, Ong, K., Jiang, H., Wu, Z., Zhuang, D., Tang, Y., Zhou, K., Zhao, J., and Liang, P. P. "ReactionBench: Evaluating Models on Fine-Grained Human Reaction Understanding from Video Stimuli." *CVPR 2026 Conference* [Under Review]

Cao, S., **Jiang, X.**, Bulusu, V., Charkrabarty, A., Hansen, M., Onat, E., Sengupta, R., and Zou, B. "Integrating flight and charging schedules in urban air mobility." In *Proceedings of the 103rd Annual Meeting of Transportation Research Board*, Washington D.C., USA, Jan **2024**. [Poster Session]

Cao, J., **Jiang, X.**, Tang, Y., Moody, J. T., Mo, Q., and Yang, H. F. "Understanding the Effect of Connector Buses on Flight Itinerary Choice." In *Proceedings of the 103rd Annual Meeting of Transportation Research Board*, Washington D.C., USA, Jan **2024**. [Oral Presentation]

Cao, J., **Jiang, X.**, Tang, Y., Mo, Q., and Yang, H. "Understanding Different Connector Buses' Effect on People's Choices in Selecting Itinerary." *2023 INFORMS Annual Meeting*, Phoenix, AZ, USA[Invited Oral Presentation]

Tsai, M., Liu, C., Yang, H., **Jiang, X.**, Zhu, M., and Wang, Y. "Unified Framework for Multi-Contrastive Learning in Spatial-Temporal Traffic Forecasting." In *Proceedings of the 103rd Annual Meeting of Transportation Research Board*, Washington D.C., USA, Jan **2024**. [Conference paper]

Bachan, J., Ye, J., **Jiang, X.**, Nguyen, T., Natarajan, M., Bremer, M., and Chan, C. "Devastator: A Scalable Parallel Discrete Event Simulation Framework for Modern C++." Invited Submission to *Proceedings of the 38th ACM SIGSIM Conference on Principles of Advanced Discrete Simulation (SIGSIM PADS '24)*, Atlanta, GA, USA, Jan **2024**. [Conference paper]

Yang, H., Zheng, W., Cai, J., Wang, P., **Jiang, X.**, Du, S., Wang, Y., and Wang, Z. "Integrating the Traffic Science with Deep Learning for City-wide Network Congestion Prediction." In *Proceedings of the 102nd Annual Meeting of Transportation Research Board*, Washington D.C., USA, Jan **2023**. Accepted. [Conference paper]

Pham, H., **Jiang, X**., & Zhang, C. (**2022**). Causality and Advanced Models in Trip Mode Prediction: Interest in Choosing Swissmetro. *2022 International Conference on Transportation and Development (ICTD)*, American Society of Civil Engineers (ASCE). [Conference paper]

Huang, W., Yan, C., & **Jiang, X**. (**2019**). Chemical and rheology evaluation on the field short-term aging of high content polymer modified asphalt (No. 19-00486). In *Proceedings of the 98th Annual Meeting of Transportation Research Board*, Washington D.C., USA, Jan **2019**. [Conference paper] URL: https://trid.trb.org/view/1572334

Journals/Conferences Peer Reviewer

TRB2023 (19 articles - Publons) ASCE ICTD 2024 Proceedings

Frontiers in Psychology (1 article - Accepted) ITSC 2023, 2024 (3 articles) AAAI 2024 (2 articles)

IEEE Intelligent Transportation Systems Society Conference 2022 (3 articles) Scientific Reports (1 article - Publons)

ASCE INTERNATIONAL CONFERENCE ON TRANSPORTATION & DEVELOPMENT 2024 (1 article)

Frontiers in Built Environment Editor, AAAI 2025, ICLR 2025, NeurIPS 2025, CVPR 2026, ICML 2026, IEEE transactions on intelligent transportation

Awards

2025 Google Software Engineering Evolution and Transformation (Sweety) Award

2024 ICRAT Best Paper Award

2023 ASCE ICTD AI in Transportation Committee Outstanding Session Organizer

2022 NSF AI workshop Phase II Travel Award

2021 Joseph M Sussman Best Paper Prize

Outstanding Student Leader, Tongji University

Shanghai Municipal Innovation & Entrepreneurship Project Award

Gold Award, 6th China International "Internet+" College Students' Innovation and Entrepreneurship Competition

Technical Skills

Languages: C/C++, CUDA, Java, Go, Python, JavaScript, SQL, TypeScript
Frameworks & Libraries: Spring Boot, Spring Cloud, gRPC, Node.js, React, MyBatis / MyBatis Plus, JWT, Shiro
Databases & Caching: MySQL, MongoDB, Redis, ElasticSearch
Cloud & Infrastructure: AWS (EC2, S3, Elastic Beanstalk), Docker, Nginx, HAProxy, gcloud
Developer Tools: Git, GitHub, Maven, Gradle, Swagger