JUNZE LIU

(217) 305-1016 ⋄ Mountain View, CA 94040
junzel1@uci.edu ⋄ junzeliu.github.io

OBJECTIVE

PhD in Computer Science with deep expertise in Large Language Models (LLMs), specializing in post-training techniques including Supervised Fine-Tuning (SFT) and Reinforcement Learning (RL). Passionate about building reliable and efficient language systems at scale.

EDUCATION

University of California, Irvine Sept. 2019 - Dec. 2024
Ph.D. in Computer Science *GPA: 4.0/4.0*

· Courses: Artificial Intelligence in Biology and Medicine, Introduction to Machine Learning, Statistical Natural Language Processing

University of Illinois at Urbana-Champaign Aug. 2016 - May. 2018
M.Eng. in Computer Engineering *GPA: 3.7/4.0*

· Courses: Computer Vision, Artificial Intelligence, Pattern Recognition, Parallel Programming, Distributed System, Interactive Computer Graphic, Virtual Reality

Shandong University Sept. 2012 - June 2016
B.E. in Communication Engineering *GPA: 86.6/100*

· Courses: Data Structure & Database Technologies, Digital Image Processing, Digital Signal Processing

SKILLS

Programming	Python, PySpark, C/C++, CUDA, Java
LLM Frameworks	PyTorch, PyTorch Lightning, vLLM, DeepSpeed, SGLang, Verl
Tools & Platforms	AWS (S3, EC2, EMR), Git, Docker, Conda, Matlab

WORK EXPERIENCE

Applied Scientist Jan 2025 - Present
Amazon *Palo Alto, CA*

· Led post-training efforts (SFT and RLHF) for both small and large LLMs (1B–32B), built on in-house and open-source models, to enhance shopping behavior understanding.
· Curated domain-specific data and designed instruction-tuning pipelines, improving performance across public benchmarks (e.g., MMLU, IFEval) and internal task-specific benchmarks by up to 15%.
· Deployed optimized models in production, collaborating with multiple teams to deliver LLM-powered features with multi-million dollar impact.

Applied Scientist Intern June 2024 - Sept 2024
Amazon *Seattle, WA*

· Advanced a 1-billion parameter foundational sequential model by integrating temporal and heterogeneous graph embeddings, significantly boosting fraud detection precision.
· Developed a novel cross-attention mechanism to fuse temporal graph and foundational model embeddings, achieving superior AUC performance over traditional tree-based methods.

Applied Scientist Intern June 2023 - Sept 2023
Amazon *Seattle, WA*

· Developed unsupervised causal discovery algorithms to enhance model explainability, feature selection, and buyer risk prediction by uncovering structural causal models within features.

· Improved cross-domain generalization of supervised classification through causal structure learning, achieving a 12% increase in AUC.

Research Assistant Jan 2019 - May 2019
UIUC *Urbana, IL*

· Developed and evaluated computer vision models for action understanding and video scene recognition, achieving state-of-the-art accuracy using ConvLSTM architectures and decision layers.
· Built an adversarial scenery dataset by web-crawling synthesized images and enhanced cross-domain robustness through weakly supervised classification techniques.

Research Assistant July 2018 - Dec 2018
Jcube *Syosset, NY*

· Designed a ConvLSTM-based action understanding algorithm for pose interaction anomaly detection.
· Improved state-of-the-art video scene recognition accuracy and cross-domain robustness by introducing decision layers and weakly supervised classification methods.

Technology Intern June 2017 - Aug 2017
Ekistic Ventures *Chicago, IL*

· Implemented a Region CNN model for illegal items recognition, adjusted to variant surveillance videos.
· Decreased false positive rate by adding multi-phase architecture to predict body-item relations.
· Designed a region labelling interface, and built a real-time tool, which is an end-to-end pipeline accepting video streams and outputting recognition results.

RESEARCH EXPERIENCE

3D Surgical Tool Detection for Retinal Eye Surgeries Ongoing
University of California, Irvine *Prof. Pierre Baldi*

· Introduced the first dataset from videos recorded during eye surgeries for surgical tool detection and classification in eye surgeries in three-dimensional space.
· Developed multiple models based on state-of-the-art detection algorithms for the tool tip detection and tool depth classification to facilitate computer-assisted interventions and surgical skill assessment.

Deep-Learning-based Neutrino Kinematic Reconstruction Ongoing
University of California, Irvine *Prof. Pierre Baldi*

· Developed 2D and 3D CNN-based models for energy and direction regression of leptons (electron and muon) produced from neutrino interactions to reconstruct the neutrino momentum.
· Implemented Sub-manifold Sparse CNN models to improve the computation and memory efficiency when processing high-dimensional data.

Geometry-aware Sparse Generative Model for Multi-layer Calorimeter Shower Simulation
Feb 2022 - Nov 2023
University of California, Irvine *Prof. Pierre Baldi*

· Built autoregressive models using machine learning algorithms to provide fast and high-quality calorimeter shower simulated data generation for high-energy physics.
· Designed a framework allowing the autoregressive model to adaptively generate multi-layer calorimeter data of different geometries.

SELECTED PUBLICATIONS

Self-Training Diffusion Language Models via Early-Step Token Distribution Calibration
Huaisheng Zhu, Ethen Liu, **Junze Liu**, Zhen Ge, Tian Wang, Jiri Gesi, Dakuo Wang, Houyu Zhang, Yufan Guo, Xian Li, Bing Yin, Sujay Sanghavi
Submitted

Shop-r1: Rewarding LLMs to Simulate Human Behavior in Online Shopping via Reinforcement Learning
Yimeng Zhang, Tian Wang, **Junze Liu**, et al.
Submitted

Nuclear Fusion Diamond Polishing Dataset
Antonios Alexos, **Junze Liu**, Shashank Galla, Sean Hayes, Kshitij Bhardwaj, Alexander Schwartz, Monika Biener, Pierre Baldi, Satish Bukkapatnam, Suhas Bhandarkar
NeurIPS, Vancouver, Canada, Dec 9, 2024

Generalizing to new geometries with Geometry-Aware Autoregressive Models (GAAMs) for fast calorimeter simulation
Junze Liu, Aishik Ghosh, Dylan Smith, Pierre Baldi, Daniel Whiteson
Journal of Instrumentation, 2023

Vitreoretinal surgical instrument tracking in 3-Dimensions using Deep Learning
(* indicates equal contribution)
Junze Liu*, Sherif Abdelkarim*, Marialejandra Diaz Ibarra*, Josiah K. To*, Pierre F. Baldi, Anjali Herekar, Baruch D. Kuppermann, Andrew W. Browne
Translational Vision Science & Technology, 2023

Deep-Learning-Based Kinematic Reconstruction for DUNE
Junze Liu, Jordan Ott, Julian Collado, Benjamin Jargowsky, Wenjie Wu, Jianming Bian, Pierre Baldi
NeurIPS Workshop on Machine Learning and the Physical Sciences, Vancouver, Canada, Dec 11, 2020

SELECTED AWARDS / SCHOLARSHIPS

- CS Travel Grants from Donald Bren School of Information and Computer Sciences 2022-2024
- Dean's Awards from Donald Bren School of Information and Computer Sciences 2019
- The Second Prize Scholarship, Shandong University 2013 2014
- Outstanding Overseas Student with Scholarship, Shandong University Dec 2013