

ZIHAN CHEN, Ph.D.

Bellevue, WA | zihanchen1995@gmail.com | 201-680-1738 | www.zihanchen.net

SUMMARY

Core scientist for the Amazon Return Conversational Agent, an LLM system handling millions of conversations per day, and lead for its global expansion. Building an internal LLM agent used daily across science, business, and engineering. Published in top information systems, FinTech, and operations management venues (POMS, ICIS).

Areas of Expertise: Large Language Models (LLMs), Graph Representation Learning (GNNs), Natural Language Processing (NLP), Deep Learning, Social Network Analysis, Time Series Analysis, Statistics, and Econometric Modeling.

WORK EXPERIENCE

Amazon.com, Inc.

Bellevue, WA

Applied Scientist

Mar 2025 – Present

- Core scientist and developer for the Amazon Return Conversational Agent, a production LLM system handling millions of customer return requests per day.
- Lead its expansion into global marketplaces: multilingual guardrails, localized prompt engineering, fine-tuning (SFT, DPO), and an automated evaluation pipeline built on human feedback.
- Build an internal LLM agent (harness, memory, hooks, MCP, RAG, UI) that automates prompt optimization and annotation job management; used daily across science, business, engineering, and data annotation teams.
- Deliver LLM agents end to end on AWS: UI, model serving (Bedrock, SageMaker), deployment (Lambda, CDK), storage (S3, DynamoDB), monitoring (CloudWatch), and access management (IAM).

Amazon.com, Inc.

Bellevue, WA

Applied Scientist Intern

Sep 2024 – Dec 2024

- Proposed a new contrastive loss for Transformer-based image matching; benchmarked against 50+ state-of-the-art baselines, improving every matching metric while reducing inference latency.
- Accepted and presented at the 2025 Amazon Machine Learning Conference.

AllianceBernstein Holding L.P.

New York, NY

Research Fellow

Jun 2024 – Aug 2024

- Customized LLMs and Transformers to extract inter-company relationships from financial text, and built graph-based deep learning models that turn them into a covariance proxy for portfolio construction.
- Raised the annualized Sharpe ratio of the Global Minimum Variance portfolio by 44% while cutting annualized volatility by 4.7% and maximum drawdown by 21% versus the strongest baseline.
- Presented at multiple industry seminars; accepted for publication in a top-tier portfolio management journal.

Jefferies Group LLC

New York, NY

Research Intern (with Data Science Doctoral Fellowship)

Mar 2022 – Jan 2023

- Used LLMs and Python (Regex, spaCy) to link millions of unlabeled financial news articles to companies and 10-K/10-Q filings; built a pipeline that extracts sentiment and text embeddings, joins them with market data (price, volume, volatility), and calibrates against market indices.
- Built deep learning models to predict stock trajectories, backtested them with Backtrader, identified the most predictive features, and analyzed inter-entity correlations for business insights.

Stevens Institute of Technology

Hoboken, NJ

Research Assistant

2019 – 2025

- Conducted research on theory-driven, explainable deep learning (graph neural networks, LLMs) for business analytics, published in FinTech, operations management (OM), and information systems (IS) journals and conferences.

Course Instructor, 3-credit Master's core courses (student evaluation: 5.0/5.0)

2023 – 2024

- Courses taught: BIA-660: Web Mining and Text Analysis (Spring 2024); BIA-580: Mathematical Foundations of Business Analytics (Fall 2023); BIA-810: Special Topics in Business Intelligence (Spring 2023).

Teaching Assistant

2018 – 2019

- Courses supported: MIS-630: Data and Knowledge Management; MIS-635: Designing the Knowledge Organization; MIS-636: Data Warehousing and Business Intelligence; BIA-652: Multivariate Data Analysis; BIA-658: Social Network Analysis.

EDUCATION

Stevens Institute of Technology Ph.D. in Data Science (STEM Major) - Concentration: Information Systems and Analytics. - Provost Doctoral Fellowship Award; Jefferies Data Science Fellowship.	Hoboken, NJ Feb 2025
Stevens Institute of Technology M.S. in Business Intelligence and Analytics Concentration: Data Science. GPA: 4.0/4.0.	Hoboken, NJ May 2019
Shandong University B.A. in Economics	Shandong, China Jun 2016

SELECTED RESEARCH PUBLICATIONS

** First/co-first author.*

Let Clickstream Talk: A Graph Neural Network Approach to Sales Forecasting - Accepted at Production and Operations Management Journal (UTD24, FT50, ABS4). - Presented at the 2022 INFORMS Annual Meeting (INFORMS), Indianapolis, US. - Best Paper Award Nominee, 2021 Pre-ICIS SIGDSA Symposium, Austin, US.	2026
Portfolio Optimization with Covariance from News-Derived Information Networks* Accepted at The Journal of Financial Data Science.	2026
Beyond Accuracy: How Procedural Traces Shift the Decision Criterion of LLM Overseers* Accepted at the Hawaii International Conference on System Sciences (HICSS), Hawaii, US.	2026
FundaPod: A Multi-Persona Agent Pod Architecture with Knowledge Graph Memory for AI-Assisted Fundamental Investment Research Accepted at the Hawaii International Conference on System Sciences (HICSS), Hawaii, US.	2026
Image-Based Item Matching on Online Return Center* Accepted at the Amazon Machine Learning Conference (AMLC, internal conference), Seattle, US.	2025
To Automate or Not? How Open Collaboration Communities Decide on Bot Approval Accepted and presented at the International Conference on Information Systems (ICIS), Bangkok, TH.	2024
Stand for Something or Fall for Everything: Predicting Misinformation Spread with Stance-Aware Graph Neural Networks* Accepted and presented at the International Conference on Information Systems (ICIS), Hyderabad, IN.	2023
Inverse Demand Function with Dual Neural Networks* - Accepted for oral presentation at the ACM International Conference on AI in Finance (ICAIF), NY, US. - Presented at the 2023 Fall Eastern Sectional Meeting, American Mathematical Society (AMS), NY, US.	2023
ChatGPT Informed Graph Neural Network for Stock Movement Prediction* Accepted at the SIGKDD Workshop on Robust NLP for Finance, Long Beach, CA, US.	2023
Making Sense of Bot Moderation in Online Communities with Large Language Models* Accepted and presented at the INFORMS Annual Meeting (INFORMS), Phoenix, Arizona, US.	2023
Does Sleep Deprivation Cause Online Incivility? Evidence from a Natural Experiment Accepted and presented at the International Conference on Information Systems (ICIS), Munich, DE.	2019

AWARDS AND GRANTS

Amazon Leadership Principles (Customer Obsession) Award — Amazon.com, Inc.	2026
Provost Doctoral Fellowship — School of Business, Stevens Institute of Technology	2019–2024
Data Science Doctoral Fellowship — Jefferies Financial Group	2022–2023
Ph.D. Student Conference Fellowship — School of Business, Stevens Institute of Technology	2020–2023
Graduate Conference Funding — Office of Graduate Academics, Stevens Institute of Technology	2023

PROFESSIONAL SERVICE

- Journal Reviewer:** IEEE Transactions on Knowledge and Data Engineering; IEEE Computational Intelligence Magazine; Technology in Society; Expert Systems with Applications; Heliyon Journal; Decision Analytics Journal; International Journal of Climatology.
- Conference Reviewer:** International Conference on Information Systems (ICIS); Americas Conference on Information Systems (AMCIS); Southern Association for Information Systems (SAIS); IEEE Symposium on Computational Intelligence for Financial Engineering and Economics (CIFER); British Machine Vision Conference Workshop on Robust Vision in Synthetic Environments (RVSSE); Hawaii International Conference on System Sciences (HICSS).
- Program Committee:** INFORMS Workshop on Data Science.

INVITED TALKS

- Incorporating LLMs and Graphs for Portfolio Construction** Jun 2024
Q-Group Investment Webinar, AllianceBernstein L.P., New York, NY, US.
- From Text to Treasure: Leveraging LLMs and Graphs for Stock Movement Predictions** Apr 2024
NLP and Machine Learning in Investment Management Conference, New York, NY, US.
- Generative AI for Business Analytics: New Paradigm, New Challenges** Oct 2023
2023 INFORMS Annual Meeting, Phoenix, Arizona, US.
- Implications of Large Language Models for the Stock Market** Oct 2023
U.S. Bank AI Research Seminar, Remote, US.
- Making Sense of AI** Oct 2022
2022 INFORMS Annual Meeting, Indianapolis, US.
- Analytics and AI for a Sustainable and Resilient Future** Dec 2021
2021 Pre-ICIS SIGDSA Symposium, Austin, US.
- The Dark Side of Social Media** Dec 2019
2019 International Conference on Information Systems, Munich, Germany.

TECHNICAL SKILLS

- Programming:** Python, PyTorch Geometric (PyG contributor), PyTorch, TensorFlow, Keras, R, Java, JavaScript, MySQL, PostgreSQL, MATLAB, HTML, SAS, Excel VBA, SPSS, Stata
- LLM Stack:** LangChain, LangGraph, LlamaIndex, Hugging Face TRL/PEFT, vLLM, DeepSpeed, verl, Ray, FAISS
- LLM Methods:** Agent harness design, hooks, tool use and function calling, Model Context Protocol (MCP), multi-agent orchestration, memory and context engineering, retrieval-augmented generation (RAG), guardrails, agent evaluation; supervised fine-tuning (SFT), parameter-efficient fine-tuning (LoRA), direct preference optimization (DPO), reinforcement learning from human feedback (RLHF), knowledge distillation, on-policy distillation (OPD)
- AWS:** Bedrock, SageMaker, Lambda, CDK, S3, EC2, DynamoDB, CloudWatch, API Gateway, IAM