

Zhuofan Jia

+1(217) 819-1162 ◇ Durham, NC ◇ zhuofan.jia@duke.edu ◇ linkedin.com/in/zhuofanjia/

Education

Duke University

Ph.D. in Computer Science, advised by Prof. Jian Pei

Aug 2023 – Expected May 2028

GPA: 4.0/4.0

University of Illinois Urbana-Champaign (UIUC)

Bachelor of Science in Computer Science and Mathematics (Dual Degree), Highest Honors

Aug 2019 – Aug 2023

GPA: 4.0/4.0

Research Interests: Large Language Models, Model Evaluation & Benchmarking, Conversational AI, Data Privacy and Unlearning, Uncertainty-Aware Data Valuation, Responsible AI

Publications

1. Minxing Zhang, Yi Yang, **Zhuofan Jia**, Xuan Yang, Jian Pei, Yuchen Zang, Xingwang Deng, Xianglong Chen. *MPCEVAL: A Benchmark for Multi-Party Conversation Generation*. Under Review KDD 2026. arXiv preprint arXiv:2603.04969.
2. Minxing Zhang, **Zhuofan Jia**, Aritra Guha, Yaron Kanza, Jian Pei, Divesh Srivastava. *Scalable Private Information Management in Large Language Models*. Under Review VLDB 2026.
3. **Zhuofan Jia** and Jian Pei. *Shapley Value on Uncertain Data*. Under Review TKDE. arXiv preprint arXiv:2601.14543.
4. Chenkai Wang, **Zhuofan Jia**, Hadjer Benkraouda, Cody Zevnik, Nicholas Heuermann, Roopa Foulger, Jonathan A. Handler, Gang Wang. *VeriSMS: A Message Verification System for Inclusive Patient Outreach against Phishing Attacks*. In Proceedings of CHI 2024.
5. Jeff Erickson, Jason Xia, Eliot Wong Robson, **Zhuofan Jia**, Tue Do, Aidan Tzur Glickman, Eric Jin, Jiwon Lee, Patrick Lin, Steven Pan, Samuel Ruggerio, Tomoko Sakurayama, Andrew Yin, Yael Gertner, Brad Solomon. *Auto-graded Scaffolding Exercises for Theoretical Computer Science*. In ASEE Annual Conference 2023.

Research Experience

Evaluating Communication Failures in Multi-Agent LLM Systems, Duke

Feb 2026 – Present

- Build evaluation frameworks to analyze communication failures in multi-agent LLM systems
- Propose local-global evaluation separating agent-level errors from system-level coordination failures
- Model agent interactions as dynamic graphs to analyze information flow, agent selection, and coordination
- Design interpretable metrics (coherence, consistency, efficiency, diversity) for failure diagnosis and supervision
- Develop automated detection methods to identify early-stage breakdowns for improving agent policies

Hierarchical Topic Tree Generation for Long-Context LLMs, Duke

Feb 2026 – Present

- Build an LLM-based system for online summarization and next-topic prediction in long conversations
- Construct reference hierarchical topic trees to provide structured supervision for fine-tuning
- Train LLMs to generate coherent topic sequences, improving long-horizon generation and reducing topic drift
- Develop multi-level generation pipelines to enhance consistency across local and global context

MPCEVAL: A Benchmark for Multi-Party Conversation Generation, Duke and Tanka

Aug 2025 – Feb 2026

- Developed a unified benchmark for both next-turn prediction and global multi-party conversation evaluation
- Proposed task-specific metrics for speaker selection, semantic coherence, novelty, and agenda progression
- Designed embedding- and topic-based evaluators to assess thematic consistency and conversational structure
- Conducted group-level and participation analysis to characterize engagement dynamics in real conversations
- Analyzed dominance and message distribution patterns to inform fair evaluation and dataset curation

Scalable Private Information Management in Large Language Models, AT&T and Duke

May 2025 – Jan 2026

- Proposed a kNN-LM architecture separating public training from private retrieval to support user-level data deletion
- Designed forgetness-enabling embedding perturbations to prevent private information leakage after removal
- Achieved 0% privacy violations with 17.5% QA accuracy gains and up to 20× perplexity reduction over baselines
- Built and evaluated private embedding databases across 200 user profiles to study privacy-utility tradeoffs
- Validated forgetness and accuracy stability via statistical and ablation analyses under varied removal rates

Shapley Value on Uncertain Data, Duke, advised by Prof. Jian Pei

Feb 2024 – Oct 2025

- Proposed an uncertainty-aware Shapley framework for fair data attribution over probabilistic data distributions
- Characterized expected contribution and contribution stability via Shapley value expectation and variance analysis
- Derived unbiased Monte Carlo estimators for expectation and variance with guarantees under sampling noise
- Developed pooled and variance-adaptive stratified estimators to reduce sampling cost and estimation error
- Achieved up to 95% variance reduction and 97% fewer samples via adaptive Monte Carlo allocation algorithms

VeriSMS, UIUC, advised by Prof. Gang Wang

Jun 2022 – May 2023

- Designed an inclusive SMS-based message verification system to mitigate phishing across diverse user abilities
- Developed a hybrid scheme combining per-message IDs and user-specific secrets to balance security and usability
- Implemented a voice-agent verification workflow requiring no apps or additional hardware for accessibility
- Conducted human-subject studies (15 exploratory, 35 validation) to evaluate understanding, trust, and usability
- Demonstrated high phishing detection precision (83%) and strong usability (SUS $\approx$ 79–85) in controlled study

Projects

Fine-Tuning and Evaluation of Small LLMs, *Large Language Models*

Oct 2025 – Dec 2025

- Fine-tuned Gemma-3-270M using SFT, LoRA/QLoRA, and curriculum training for reasoning and factual QA tasks
- Implemented DPO/RLHF, self-correction, chain-of-thought, and RAG for improved generalization
- Integrated knowledge distillation from Gemma3-1B-IT and experiment with data synthesis for domain adaptation
- Achieved $3.2\times$ gain on TriviaQA exact match and $1.9\times$ gain on ARC-C accuracy over the zero-shot baseline

Benchmarking LLM Agents on False Hypothesis Rejection, *Large Language Model Agents*

Oct 2025 – Dec 2025

- Built a benchmark to test LLM agents' ability to reject, refine, or confirm hypotheses from real datasets
- Generated incrementally false hypotheses to measure agents' sensitivity and specificity in falsification tasks
- Evaluated reasoning pipelines (ReAct, CoT, POPPER) on rejection accuracy, efficiency, and bias
- Proposed an auditable framework for assessing scientific reasoning and data-driven discovery in agentic models

Professional Experience

Software Engineer Intern, HERE Technologies – Chicago, IL

Jan 2023 – May 2023

- Led research on time-series traffic patterns at road segments to determine the operational status of stores, leveraging machine learning for pattern recognition
- Developed and tested XGBoost, neural networks, and logistic regression models on SageMaker with F1 > 95%
- Implemented data obfuscation algorithms with GeoPandas that replace precise locations with broader geographic units, such as Traffic Analysis Zones, to enhance trip privacy

Machine Learning Engineer Intern, Discovery Partners Institute – Chicago, IL

Jan 2022 – May 2023

- Developed an ML-powered medical search engine, integrating clinical data for enhanced information retrieval
- Built a scalable web crawler using BeautifulSoup to extract structured content for improved query relevance
- Implemented a content classification neural network in PyTorch, using BERT for keyword extraction
- Integrated a semantic search microservice with FastAPI, applying deep learning and NLP to rank and retrieve relevant results

Software Developer, UIUC, advised by Prof. Jeff Erickson

Aug 2022 – May 2023

- Designed and deployed innovative auto-graded practice problems for algorithms and data structures using Python, JavaScript, and React on PrairieLearn platform, enhancing learning for 3,500+ students
- Developed scaffolded exercises to reinforce theoretical computer science concepts, improving student engagement

Patents

1. Jingwei Xu, Weimin Huang, Bruce Bernhardt, Yuxin Guan, **Zhuofan Jia**, Jingbo Li. *Dwell Time Modelling for Point of Interests*. US Patent Application 19/342094, filed September 26, 2025.
2. Jingwei Xu, Weimin Huang, Bruce Bernhardt, Yuxin Guan, **Zhuofan Jia**, Jingbo Li. *Trip Data Privacy Based on Obfuscation*. US Patent Application 19/313266, filed August 28, 2025.

Teaching Experience

- **CompSci 516 Database Systems**, Teaching Assistant, Duke
- **CompSci 316 Introduction to Databases**, Teaching Assistant, Duke

Spring 2025

Fall 2024

- **CS 446 Machine Learning**, Course Assistant, UIUC
 - **CS 374 Intro to Algorithms & Models of Computation**, Course Developer, UIUC
 - **CS 125 Intro to Computer Science**, Course Assistant, UIUC
- Spring 2023

Fall 2022 – Spring 2023

Spring 2020 – Fall 2020

Professional Services

- **Reviewer** for ACM Transactions on Knowledge Discovery from Data (TKDD), 2025

Honors and Awards

- Duke CS PhD Fellowship, Duke, 2023-24
- University Honor – The Bronze Tablet, UIUC, 2023
- C.W. Gear Outstanding Undergraduate Award, UIUC, 2022
- Dean’s List, UIUC, 2019-23
- James Scholar, UIUC, 2019-23

Skills

Programming Languages:	Python, C++ , Java, JavaScript, TypeScript, SQL, Bash
LLM Training & Techniques:	SFT, RLHF, DPO, LoRA, QLoRA, RAG
Machine Learning & Data Tools:	PyTorch, Transformers, vLLM, pandas, NumPy, Spark, scikit-learn, XGBoost
Full-Stack & Data Systems:	AWS, FastAPI, Flask, React.js, Node.js, MySQL, PostgreSQL, MongoDB, Neo4j