Dongkuan (DK) Xu **Assistant Professor of Computer Science**

CONTACT INFORMATION

Web: dongkuanx27.github.io/ *Google Scholar:* [Link]
E-mail: dxu27@ncsu.edu *Twitter:* https://twitter.com/DongkuanXu
Mobile: 814-699-0860 *Updated:* Sep 9th, 2024

RESEARCH INTERESTS

My research is fundamentally grounded in advancing **Artificial General Intelligence**, with particular emphasis on studying the autonomy of intelligent agents (*task planning, external tool use*), decision/reasoning reliability (*alignment, uncertainty, adaptability*), and resource efficiency (*parameter, data, computation*) in generative AI systems (*ChatGPT, GPT-X, diffusion models*). I'm leading the **NCSU Generative Intelligent Computing Lab**. My long-term research goal is to liberate AI productivity and democratize its application to serve a broader range of populations and real-world applications, equally, sustainably, and responsibly.

WORKING

Assistant Professor, North Carolina State University, NC, USA. Aug 2022-Present
- Department of Computer Science
- **Microsoft Accelerating Foundation Models Research Award**, 2024
- **NCSU Carla Savage Award**, 2024
- **ICCCN Best Paper Award**, 2023

EDUCATION

PhD, Penn State University, PA, USA. 2022
- College of IST Award for Excellent Teaching [Top 2]

MS, University of Chinese Academy of Sciences, Beijing, China 2017
- Chinese Academy of Sciences President's Fellowship [Top 1]

BE, Renmin University of China, Beijing, China 2014

PUBLICATION SUMMARY

Published: 61 papers, **30** first/advising-authored papers, and **10** filed patents.
Impact: 3773 citations, h-index: **17**, i10-index: **26** (as of Sep 9th, 2024). My publications can be generally categorized as follows (with a representative paper under each category).
Published at NeurIPS, ICLR, AAAI, ACL, EMNLP, NAACL, CVPR, ECCV, ICCV, DAC, etc.

- **Improving Parameter Efficiency in Foundation Models** [17][21][35][38][40][43][44][49]

  — X. Liu, B. Lei, R. Zhang, **D. Xu**. Adaptive Draft-Verification for Efficient Large Language Model Decoding [C]. **(ArXiv, Aug 2024)** Project [Link], Demo [Link], Paper [Link]

- **Improving Computation Efficiency of Foundation Models** [14][29][30][36][37][2]

  — S. Tang, Y. Wang, C. Ding, Y. Liang, Y. Li, **D. Xu**. AdaDiff: Accelerating Diffusion Models through Step-Wise Adaptive Computation [C]. The 18th European Conference on Computer Vision **(ECCV 2024)**

- **Improving Data Efficiency in Foundation Models** [18][13][39][45][47][51][52][54]

  — L. Zhang, J. Zhang, B. Lei, S. Mukherjee, X. Pan, B. Zhao, C. Ding, Y. Li, **D. Xu**. Accelerating Dataset Distillation via Model Augmentation [C]. The 34th IEEE/CVF Conference on Computer Vision and Pattern Recognition **(CVPR 2023, Highlight Paper, 2.5%)**

CURRENT PROJECTS

Harnessing Links between Historical Business & Household Microdata and Street-View Images to Assess Transit-Induced Neighborhood Changes at Small Spatial Scales
- Source of Support: NSF
- Project Start Date: 08/2024
- Project End Date: 08/2027

- PI: Eleni Bardaka (NC State, Civil Engineering)
- Co-PI: DK Xu (NC State, Computer Science)
- Total Project Amount: $396,600

Collaborative Research: CyberTraining: Implementation: Medium: EcoTern: Pioneering a CI Workforce for Sustainable and Transdisciplinary Environmental Science Research

- Source of Support: NSF
- Project Start Date: 12/2024
- Project End Date: 11/2027
- NCSU Team: Xipeng Shen (PI, Systems), DK Xu (Co-PI, AI), Roy He (Co-PI, Ocean)
- FIU Team: Wenqian Dong (PI, HPC), J. Obey (Co-PI, Climate), J. Liu (Co-PI, Simulation)
- Total Project Amount: $979,901

Scalable and Adaptable Evaluation of LLMs' Trustworthiness Through Generative Techniques

- Source of Support: Microsoft Accelerating Foundation Models Research
- Project Start Date: 01/2024
- Project End Date: 06/2025
- Single PI: DK Xu (NC State, Computer Science)
- Total Project Amount: $50,000

Community Engagement

- Workshop on DL-Hardware Co-Design for Generative AI Acceleration @DAC'24, Chair
- Workshop on Dataset Distillation for Computer Vision @CVPR'24, Co-Chair
- 2nd Resource-efficient Learning for Knowledge Discovery Workshop @KDD'24, Co-Chair
- 1st Workshop on DL-Hardware Co-Design for AI Acceleration @AAAI'23, Chair
- 1st Resource-Efficient Learning for Knowledge Discovery Workshop @KDD'23, Co-Chair
- ML & NLP Learning Community (Chinese), Founding Committee Member
- ACM SIGAI Newsletter, Column Editor
- NSF CAREER Panel Reviewer, 2023
- Area Chair, Session Chair, (Senior) Program Committee Member for $\geq$ 50 times

Education Outreach

- Grand Challenges Scholars Program (GCSP-REU) Summer 2024, Research Mentor
- NCSU Educational Workshops (Integrating ChatGPT into K-12 Classrooms), Co-Chair
- NSF REU Site (SRCA, Socially Relevant Computing and Analytics), UG Mentor
- NCSU CSC 298 (Introduction to Computer Science Research Methods), Mentor (2 UGs)
- NCSU COE REU Program, UG Mentor (**$3,000 Research Award awarded to my student**)
- NSF-funded IUSE project (ExplainIt), UG Class Instructor
- NSF REU Site proposal (Topic: Algorithms and Theory), Mentor (2 UGs for 3-5 years)

Open-source Project

Gentopia.AI: A Collaborative Platform for Tool-Augmented LLMs

- **Goal:** Aim to specialize & share agents to overlay collective growth for greater intelligence
- **Teams:** Researchers from NC State, George Mason, NYU, UMich, CMU
- **Web:** https://github.com/Gentopia-AI, **Demo** [link], **Quick Start** [link]
- **Paper:** Accepted to EMNLP'23 (System Demo) [link]

Peer-reviewed conference and journal papers

[1] H. Reichert, B. Tabarsi, Z. ZHang, C. Fennell, I. Bhandari, D. Robinson, M. Drayton, C. Crofton, M. Lococo, **D. Xu**, T. Barnes. Empowering Secondary School Teachers: Creating, Executing, and Evaluating a Transformative Professional Development Course on ChatGPT [C]. IEEE Frontiers in Education Conference 2024 **(FIE'24)**

[2] S. Tang, Y. Wang, C. Ding, Y. Liang, Y. Li, and **D. Xu**. AdaDiff: Accelerating Diffusion Models through Step-Wise Adaptive Computation [C]. The 18th European Conference

on Computer Vision **(ECCV'24)**

[3] J. Liu, Z. Peng, **D. Xu**, Y. Liu. Revolutionizing Wireless Modeling and Simulation with Network-Oriented LLMs [C]. The 43rd IEEE International Performance Computing and Communications Conference **(IPCCC'24)**

[4] Z. Zhang, Y. Liu, Z. Peng, M. Chen, **D. Xu**, and S. Cui. Digital Twin-Assisted Data-Driven Optimization for Reliable Edge Caching in Wireless Networks [J]. IEEE Journal on Selected Areas in Communications **(IEEE JSAC, 2024, Impact Factor is 16.4)**

[5] P. Dong, J. Zhuang, Z. Yang, S. Ji, Y. Li, **D. Xu**, H. Huang, J. Hu, A. Jones, Y. Shi, Y. Wang, P. Zhou. EQ-ViT: Algorithm-Hardware Co-Design for End-to-End Acceleration of Real-Time Vision Transformer Inference on Versal ACAP Architecture [C]. The International Conference on Hardware/Software Codesign and System Synthesis **(CODES+ISSS'24)**

[6] X. Wang, S. Duan, X. Yi, J. Yao, S. Zhou, Z. Wei, P. Zhang, **D. Xu**, M. Sun, X. Xie. On the Essence and Prospect: An Investigation of Alignment Approaches for Big Models [C]. International Joint Conference on Artificial Intelligence (Survey Track) **(IJCAI'24)**

[7] Y. Wang, Q. Zhao, **D. Xu**, and X. Liu. Purpose Enhanced Reasoning through Iterative Prompting: Uncover Latent Robustness of ChatGPT on Code Comprehension [C]. 2024 International Joint Conference on Artificial Intelligence **(IJCAI'24)**

[8] X. Luo, Z. Li, Z. Peng, **D. Xu**, Y. Liu. RM-Gen: Conditional Diffusion Model-Based Radio Map Generation for Wireless Networks [C]. International Federation for Information Processing Networking Conference **(IFIP/IEEE Networking'24)**

[9] B. Lei, **D. Xu**, R. Zhang, and B.K Mallick. Embracing Unknown Step by Step: Towards Reliable Sparse Training in Real World [J]. **(Transactions on Machine Learning Research, 2024)**

[10] B. Lei, **D. Xu**, R. Zhang, S. He, B. K. Mallick. Balance is Essence: Accelerating Sparse Training via Adaptive Gradient Correction [C]. The 2024 Conference on Parsimony and Learning **(CPAL'24)**

[11] Z. Zhang*, Z. Dong*, Y. Shi, N. Matsuda, T. Price, **D. Xu**. Students' Perceptions and Preferences of Generative Artificial Intelligence Feedback for Programming [C]. The 14th Symposium on Educational Advances in Artificial Intelligence **(AAAI/EAAI'24)**

[12] Z. Wang, Q. Zhao, J. Cui, X. Liu, and **D. Xu**. AutoST: High-performance and Energy-efficient Spiking Transformer Architecture Search [C]. The 2024 IEEE International Conference on Acoustics, Speech, and Signal Processing **(ICASSP'24)**

[13] L. Zhang, J. Zhang, B. Lei, S. Mukherjee, X. Pan, B. Zhao, C. Ding, Y. Li, and **D. Xu**. Accelerating Dataset Distillation via Model Augmentation [C]. The 34th IEEE/CVF Conference on Computer Vision and Pattern Recognition **(CVPR'23, Highlight Paper**) Acceptance rate: **235/9155=2.5%**

[14] S. Tang, Y. Wang, Z. Kong, T. Zhang, Y. Li, C. Ding, Y. Wang, Y. Liang, and **D. Xu**. You Need Multiple Exiting: Dynamic Early Exiting for Accelerating Unified Vision Language Model [C]. The 34th IEEE/CVF Conference on Computer Vision and Pattern Recognition **(CVPR'23)**

[15] B. Xu, X. Liu, H. Shen, Z. Han, Y. Li, M. Yue, Z. Peng, Y. Liu, Z. Yao, and **D. Xu**. Gentopia: A Collaborative Platform for Tool-Augmented LLMs [C]. The 2023 Conference on Empirical Methods in Natural Language Processing. **(EMNLP'23, System Track)**

[16] J. Li, Q. Lei, W. Cheng, and **D. Xu**. Towards Robust Pruning: An Adaptive Knowledge-Retention Pruning Strategy for Language Models [C]. The 2023 Conference on Empirical Methods in Natural Language Processing. **(EMNLP'23)**

[17] J. Li, W. Gao, Q. Lei, and **D. Xu**. Breaking through Deterministic Barriers: Randomized Pruning Mask Generation and Selection [C]. The 2023 Conference on Empirical Methods in Natural Language Processing. **(EMNLP'23, Findings)**

[18] J. Gu, Z. Nan, Z. Peng, X. Shen, and **D. Xu**. Co-evolving Data-driven and NLU-driven Synthesizers for Generating Code in Domain Growth and Data Scarcity [C]. The 2023 Conference on Empirical Methods in Natural Language Processing. **(EMNLP'23, Pan-DL Workshop)**

[19] D. Zhu, B. Lei, J. Zhang, Y. Fang, Y. Xie, R. Zhang, and **D. Xu**. Rethinking Data Distillation: Do Not Overlook Calibration [C]. International Conference on Computer Vision **(ICCV'23)**

[20] J. Wang, X. Yang, S. Cui, L. Che, L. Lyu, **D. Xu**, and F. Ma. Towards Personalized Federated Learning via Heterogeneous Model Reassembly [C]. The 37th Conference on Neural Information Processing Systems **(NeurIPS'23)**

[21] S. Li, H. Mei, J. Li, H. Wei, and **D. Xu**. Toward Efficient Traffic Signal Control: Smaller Network Can Do More [C]. 62nd IEEE Conference on Decision and Control **(CDC'23)**

[22] B. Lei, R. Zhang, **D. Xu**, and B. K Mallick. Calibrating the Rigged Lottery: Making All Tickets Reliable [C]. The 11th International Conference on Learning Representations **(ICLR'23)**

[23] Q. Zhang, S. Chen, **D. Xu**, Q. Cao, X. Chen, T. Cohn, and M. Fang. A Survey for Efficient Open Domain Question Answering [C]. The 61th Annual Meeting of the Association for Computational Linguistics **(ACL'23)**

[24] L. Wu, B. Lei, **D. Xu**, and D. Zhou. Towards Reliable Rare Category Analysis on Graphs via Individual Calibration [C]. The 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining **(KDD'23)**

[25] C. Liu, D. Doshi, M. Bhargava, R. Shang, J. Cui, **D. Xu**, and E. Gehringer. Labels Are Not Necessary: Assessing Peer-Review Helpfulness Using Domain Adaptation Based on Self-Training [C]. The 18th Workshop on Innovative Use of NLP for Building Educational Applications **(BEA'23)**

[26] Z. Dong, **D. Xu**. Exploring the Augmented Large Language Model with Mathematical tools in Personalized and Efficient Education [C]. The 6th International Conference on Artificial Intelligence and Big Data **(ICAIBD'23)**

[27] Y. Liu, M. Chen, **D. Xu**, Z. Yang, and S. Zhao. E-App: An Environment-Aware Access Point Planning Framework for mmWave Wireless LANs [C]. The 32nd International Conference on Computer Communications and Networks **(ICCCN'23, Best Paper Award)**

[28] Y. X., D. Zhu, B. Lei, **D. Xu**, and R. Zhang. Efficient Informed Proposals for Discrete Distributions via Newton's Series Approximation [C]. The 26th International Conference on Artificial Intelligence and Statistics **(AISTATS'23)**

[29] S. Huang, H. Fang, K. Mahmood, B. Lei, N. Xu, B. Lei, Y. Sun, **D. Xu**, Wu. Wen, and C. Ding. Neurogenesis Dynamics-inspired Spiking Neural Network Training Acceleration [C]. The 60th Design Automation Conference **(DAC'23)**

[30] S. Huang, B. Lei, **D. Xu**, H. Peng, Y. Sun, M. Xie, and C. Ding. Dynamic Sparse Training via Balancing the Exploration-Exploitation Trade-off [C]. The 60th Design Automation Conference **(DAC'23)**

[31] J. Li, T. Zhang, E. Yan, and **D. Xu**. FP8-BERT: Post-Training Quantization for Transformer [C]. The 1st Workshop on DL-Hardware Co-Design for AI Acceleration **(DCAA'23)**

[32] Y. Xie, Z. Li, H. Bao, X. Jia, **D. Xu**, X. Zhou, and S. Skakun. Auto-CAM: Label-Free Earth Observation Imagery Composition and Masking Using Spatio-Temporal Dynamics [C]. The 37th AAAI International Conference on Artificial Intelligence **(AAAI'23)**

[33] D. Luo, W. Cheng, Y. Wang, **D. Xu**, J. Ni, W. Yu, X. Zhang, Y. Liu, Y. Chen, H. Chen, and X. Zhang. Time Series Contrastive Learning with Information-Aware Augmentations [C]. The 37th AAAI International Conference on Artificial Intelligence **(AAAI'23)**

[34] Y. Tian, Weizhi Gao, Qin Zhang, Pu Sun, and **D. Xu**. Improving long-tailed classification by disentangled variance transfer [J]. **Internet of Things (2023)**: 100687.

[35] **D. Xu**, S. Mukherjee, X. Liu, D. Dey, W. Wang, X. Zhang, A. H. Awadallah, and J. Gao. Few-shot Task-agnostic Neural Architecture Search for Distilling Large Language Models [C]. The 36th Conference on Neural Information Processing Systems **(NeurIPS'22)**

[36] I. Yen, Z. Xiao, and **D. Xu**. S4: a High-sparsity, High-performance AI Accelerator [C]. Sparsity in Neural Networks 2022 Workshop **(SNN'22)**

[37] S. Huang, N. Liu, Y. Liang, H. Peng, H. Li, **D. Xu**, M. Xie, and C. Ding. An Automatic and Efficient BERT Pruning for Edge AI Systems [C]. The 23rd IEEE International Society for Quality Electronic Design **(ISQED'22)**

[38] S. Huang*, **D. Xu***, I. Yen, S. Chang, B. Li, S. Chen, M. Xie, H. Liu, and C. Ding. Sparse Progressive Distillation: Resolving Overfitting under Pretrain-and-Finetune Paradigm [C]. The 60th Annual Meeting of the Association for Computational Linguistics **(ACL'22)**
Acceptance rate: 714/3350=21.3%

[39] **D. Xu**, W. Cheng, D. Luo, H. Chen, and X. Zhang. InfoGCL: Information-Aware Graph Contrastive Learning [C]. The 35th Conference on Neural Information Processing Systems **(NeurIPS'21)**
Acceptance rate: 2372/9122=26.0%

[40] **D. Xu**, I. Yen, J. Zhao, and Z. Xiao. Rethinking Network Pruning - under the Pre-train and Fine-tune Paradigm [C]. 2021 Annual Conference of the North American Chapter of the Association for Computational Linguistics **(NAACL-HLT'21)**
Acceptance rate: 477/1797=26.5%

[41] X. Dong, Y. Zhu, Z. Fu, **D. Xu**, and G. de Melo. Data Augmentation with Adversarial Training for Cross-Lingual NLI [C]. The 59th Annual Meeting of the Association for Computational Linguistics **(ACL'21)**
Acceptance rate: 714/3350=21.3%

[42] **D. Xu**, W. Cheng, J. Ni, D. Luo, Masanao Natsumeda, D. Song, B. Zong, H. Chen, and X. Zhang. Deep Multi-Instance Contrastive Learning with Dual Attention for Anomaly Precursor Detection [C]. The 21th SIAM International Conference on Data Mining **(SDM'21)**
Acceptance rate: 85/400=21.3%

[43] **D. Xu**, W. Cheng, X. Dong, B. Zong, W. Yu, J. Ni, D. Song, X. Zhang, H. Cheng, and X. Zhang. Multi-Task Recurrent Modular Networks [C]. The 35th AAAI International Conference on Artificial Intelligence **(AAAI'21)**
Acceptance rate: 1692/7911=21.4%

[44] **D. Xu**, J. Liang, W. Cheng, H. Wei, H. Cheng, and X. Zhang. Transformer Style Relational Reasoning with Dynamic Memory Updating for Temporal Network Modeling [C]. The 35th AAAI International Conference on Artificial Intelligence **(AAAI'21)**
Acceptance rate: 1692/7911=21.4%

[45] H. Wei, **D. Xu**, J. Liang, and Z. Li. How Do We Move: Modeling Human Movement with System Dynamics [C]. The 35th AAAI International Conference on Artificial Intelligence **(AAAI'21)**
Acceptance rate: 1692/7911=21.4%

[46] J. Liang, Y. Wu, **D. Xu**, and V. Honavar. Longitudinal Deep Kernel Gaussian Process Regression [C]. The 35th AAAI International Conference on Artificial Intelligence **(AAAI'21)**
Acceptance rate: 1692/7911=21.4%

[47] D. Luo, W. Cheng, **D. Xu**, W. Yu, B. Zong, H. Chen, and X. Zhang. Parameterized Explainer for Graph Neural Network [C]. The 34th Conference on Neural Information Processing Systems **(NeurIPS'20)**
Acceptance rate: 1900/9454=20.1%

[48] X. Dong, Y. Zhu, Y. Zhang, Z. Fu, **D. Xu**, S. Yang, and G. de Melo. Leveraging Adversarial Training in Self-Learning for Cross-Lingual Text Classification [C]. The 43rd International ACM SIGIR Conference on Research and Development in Information Retrieval **(SIGIR'20)**
Acceptance rate: 300/1062=28.2%

[49] **D. Xu**, W. Cheng, B. Zong, D. Song, J. Ni, W. Yu, Y. Liu, H. Chen, and X. Zhang. Tensorized LSTM with Adaptive Shared Memory for Learning Trends in Multivariate Time Series [C]. The 34th AAAI International Conference on Artificial Intelligence **(AAAI'20)**
Acceptance rate: 1591/7737=20.6%

[50] J. Liang, **D. Xu**, Y. Sun, and V. Honavar. Longitudinal Multi-Level Factorization Machines [C]. The 34th AAAI International Conference on Artificial Intelligence **(AAAI'20)**
Acceptance rate: 1591/7737=20.6%

[51] **D. Xu**, W. Cheng, D. Luo, X. Liu, and X. Zhang. Spatio-Temporal Attentive RNN for Node Classification in Temporal Attributed Graphs [C]. The 28th International Joint Conference on Artificial Intelligence **(IJCAI'19)**
Acceptance rate: 850/4752=17.9%

[52] **D. Xu**, W. Cheng, D. Luo, Yameng Gu, X. Liu, J. Ni, B. Zong, H. Chen, and X. Zhang. Adaptive Neural Network for Node Classification in Dynamic Networks [C]. The 19th IEEE International Conference on Data Mining **(ICDM'19)**
Acceptance rate: 183/930=19.7%

[53] **D. Xu**, W. Cheng, B. Zong, J. Ni, D. Song, W. Yu, Y. Chen, H. Chen, and X. Zhang. Deep Co-Clustering [C]. The 19th SIAM International Conference on Data Mining **(SDM'19)**
Acceptance rate: 90/397=22.7%

[54] J. Ni, S. Chang, X. Liu, W. Cheng, H. Chen, **D. Xu**, and X. Zhang. Co-Regularized Deep Multi-Network Embedding [C]. The 27th International Conference on World Wide Web **(WWW'18)**
Acceptance rate: 170/1175=14.5%

[55] Y. Tian, **D. Xu**, and C. Zhang. A Review of Multi-Instance Learning Research [J]. Operations Research Transactions, 2018, 02: 1-17

[56] **D. Xu**, J. Wu, D. Li, Y. Tian, X. Zhu, and X. Wu. SALE: Self-Adaptive LSH Encoding for Multi-Instance Learning [J]. **Pattern Recognition**, 2017 **(7.74 impact factor)**

[57] D. Li, **D. Xu**, J. Tang, and Y. Tian. Metric Learning for Multi-Instance Classification with Collapsed Bags [C]. The 30th IEEE International Joint Conference on Neural Networks **(IJCNN'17)**

[58] D. Li, W. Zhang, **D. Xu**, and Y. Tian. Multi-Metrics Classification Machine [C]. International Conference on Information Technology and Quantitative Management **(ITQM'16)**

[59] **D. Xu**, and Y. Tian. A Comprehensive Survey of Clustering Algorithms [J]. Annals of Data Science, 2015, 2(2): 165-193

[60] **D. Xu**, T. Chen, and W. Xu. A Support Vector Machine-Based Ensemble Prediction for Crude Oil Price with VECM and STEPMRS [J]. International Journal of Global Energy Issues, 2015

[61] **D. Xu**, Y. Zhang, C. Cheng, W. Xu, and L. Zhang. A Neural Network-Based Ensemble Prediction Using PMRS and ECM [C]. The 47th IEEE Hawaii International Conference on System Sciences **(HICSS'14)**

INDUSTRY EXPERIENCE

Microsoft Research (MSR), Redmond, WA 2021
- Research Intern, Mentors: Subho Mukherjee, X. Liu, D. Dey, A. H. Awadallah, J. Gao
- Project: Task-agnostic Auto-Transformer Search [NeurIPS 2022]

Moffett.AI, Los Altos, CA 2020
- Research Intern, Mentor: I. Yen, Co-founder
- Project: Data-free Model Compression [NAACL 2021 & a U.S. patent]

NEC Labs America, Princeton, NJ 2019
- Research Intern, Mentor: W. Cheng
- Project: Knowledge Transfer in Multi-Task Learning [AAAI 2021]
- Project: Trend Learning in Multivariate Time Series [AAAI 2020]

NEC Labs America, Princeton, NJ 2018
- Research Intern, Mentor: W. Cheng, Senior Researcher
- Project: Contrastive Anomaly Detection [SDM 2021]

ACADEMIA EXPERIENCE

Penn State University 2017-2022
- Graduate Research Assistant, Adviser: X. Zhang
- Thesis: Resource-efficient Deep Learning: Democratizing AI at Scale

Chinese Academy of Sciences, Beijing, China 2014-2017
- Graduate Research Assistant, Adviser: Y. Tian
- Thesis: Efficient Multi-instance Learning

Renmin University of China, Beijing, China 2012-2014
- Undergraduate Research Assistant, Adviser: W. Xu
- Thesis: Ensemble Forecasting Model for Time Series Data

TEACHING EXPERIENCE

Instructor at NC State

- CSC 422: Automated Learning and Data Analysis Spring'23, Fall'24
  Course Materials: Introduction to Data Mining (Second Edition)

- CSC 791&591: Advanced Topics in Efficient Deep Learning — Fall'22, Fall'23, Spring'24
  Course Materials: Dive into Deep Learning

Teaching Assistant at Penn State

- SRA 268, Visual Analytics — Fall 2021
  Instructor: Prof. Mahir Akgun
  Course Materials: Visual Analytics with Tableau
  (Responsible for teaching lab classes of 46 students)
- SRA 450, Cybercrime and Cyberwar — Fall 2021
  Instructor: Prof. John Hodgson
  Course Materials: Cybersecurity: What Everyone Needs to Know
- DS/CMPSC 410, Programming Models for Big Data — Spring 2021
  Instructor: Prof. John Yen
  Course Materials: Learning Spark
- SRA 365, Statistics for Security and Risk Analysis — Fall 2020
  Instructor: Dr. James Farrugia
  Course Materials: Discovering Statistics Using R
- DS 402, Introduction to Social Media Mining — Spring 2020
  Instructor: Prof. Suhang Wang
  Course Materials: Social Media Mining: An Introduction
- SRA 365, Statistics for Security and Risk Analysis — Spring 2019
  Instructor: Dr. Katherine Hamilton
  Course Materials: Foundations and Practice of Intermediate Statistics
- IST 210, Organization of Data — Fall 2018
  Instructor: Prof. X. Zhang
  Course Materials: Database Systems Concepts
  (The Award for Excellence in Teaching Support)

Guest Lecturer

- COSI 133A, Graph Mining — Fall 2021
  Brandeis University, Slides [Link]
- COSI 165B, Deep Learning — Spring 2021
  Brandeis University, Slides: [Link]

MENTORING EXPERIENCE

Ph.D. Students

- Li-Chia (Jerry) Chang, Ph.D. at NC State University
  Topic: Agentic LLM-Powered Agents
- Muhammad Alahmadi, Ph.D. at NC State University
  Topic: Data-Centric Reliability in LLMs
- Kaushik Pillalamarri, Ph.D. at NC State University
  Topic: Multi-Modal Foundation Models

Master Students

- Teddy Chen, Master at NC State University
  Topic: Implicit Bias in Large Language Models
- Homak Patel, Master at NC State University
  Topic: Reliable Retrieval-Augmented Generation for Education

Undergraduate Researchers

- Noah (Precious) Donkor, Undergraduate at NC State University
  Topic: Reliable LLM-powered Agents
- Aditya Basarkar, Undergraduate at NC State University
  Topic: LLM-powered Math Reasoning

Intern Researchers

- Xukun Liu, Master at Northwestern University
  Topic: Accelerating LLM Decoding
- Berwin Chen, Master at University of Birmingham
  Topic: Retrieval-Augmented Generation for Scientific Discovery
- Shanlin Liu, Master at University of Shanghai for Science and Technology
  Topic: Retrieval-Augmented Generation for Scientific Discovery
- Huanhuan Ma, Master at Chinese Academy of Sciences
  Topic: Trustworthiness Evaluation of LLMs
- Zihao Mao, Undergraduate at Wuhan University
  Topic: Accelerating Diffusion Models
- Feifei Qiao, Master at Tongji University
  Topic: Robust Inference in Agentic LLMs

Alumni

- Zhiyuan Peng, Postdoc, 2023-2024, now Meta
  Topic: Tool-Augmented Large Language Model
- Shengkun Tang, PhD, 2023-2023, now MBZUAI
  Topic: Multi-Modal Foundation Models
- Chengyuan Liu, PhD, 2022-2024
  Topic: Data-Efficient Large Language Models for Education
- Bowen Lei, Intern, 2022-2024, now Apple
  Topic: Theoretical Foundations of Efficient Learning
- Zhengdong Zhang, Intern, 2023-2024, now Amazon
  Topic: Reliable Large Language Models for Education
- Zihan (Z) Dong, Undergraduate, 2023-2024, now Georgia Tech
  Topic: Reliable Large Language Models for Education
- Binfeng Xu, Intern, 2023-2023, now Samsung
  Topic: Tool-Augmented Large Language Model
- Cong Zeng, Intern, 2024-2024, now MBZUAI
  Topic: LLM-Generated Content Detection
- John Zhu, Master, 2024, now NC State
  Topic: Trustworthy Evaluation of LLMs

PROFESSIONAL SERVICE

Panel Reviewer

- NSF CORE Program, 2024
- NSF CAREER Program, 2023

Column Editor

- ACM SIGAI Newsletter

Workshop Chair

- Workshop on DeepLearning-Hardware Co-Design for Generative AI Acceleration @DAC2024
- The First Workshop on Dataset Distillation for Computer Vision @CVPR2024
- The 2nd Workshop on Resource-Efficient Learning for Knowledge Discovery @KDD2024
- The 1st Workshop on DL-Hardware Co-Design for AI Acceleration @AAAI2023
- The 1st Workshop on Resource-Efficient Learning for Knowledge Discovery @KDD2023

Session Chair
- Scalable, Distributed Systems & Trustable AI @KDD2022
- Deep Learning: New Architectures and Models @KDD2022

Academic Committee Member
- Machine Learning & Natural Language Processing Community (MLNLP)

Area Chair
- The Joint International Conference on Computational Linguistics, Language Resources and Evaluation (LREC-COLING) 2024

Senior Program Committee
- AAAI Conference on Artificial Intelligence (AAAI) 2024, 2025
- International Joint Conferences on Artificial Intelligence (IJCAI) 2021

Program Committee
- Neural Information Processing Systems (NeurIPS) 2020, 2021, 2022, 2023
- International Conference on Learning Representations (ICLR) 2021, 2022, 2023, 2024
- International Conference on Machine Learning (ICML) 2021, 2022, 2023
- AAAI Conference on Artificial Intelligence (AAAI) 2020, 2021, 2022, 2023
- SIGKDD Conference on Knowledge Discovery and Data Mining (KDD) 2020-2023
- Association for Computational Linguistics (ACL) Rolling Review 2022
- North American Chapter of the Association for Computational Linguistics (NAACL) 2021
- Conference on Empirical Methods in Natural Language Processing (EMNLP) 2020, 2021
- International Conference on Computational Linguistics (COLING) 2022
- Learning on Graphs Conference (LoG) 2022
- International Joint Conferences on Artificial Intelligence (IJCAI) 2020, 2022
- ACM International Conference on Web Search and Data Mining (WSDM) 2022
- SIAM International Conference on Data Mining (SDM) 2022
- European Chapter of the Association for Computational Linguistics (EACL) 2021
- Conference on Information and Knowledge Management (CIKM) 2020, 2021, 2022
- Asia-Pacific Chapter of the Association for Computational Linguistics & International Joint Conference on Natural Language Processing (AACL-IJCNLP) 2020
- International Joint Conference on Neural Networks (IJCNN) 2018, 2019, 2020, 2021

Journal Reviewer
- IEEE Transactions on Pattern Analysis and Machine Intelligence (TPAMI)
- IEEE Transactions on Neural Networks and Learning Systems (TNNLS)
- IEEE Transactions on Knowledge and Data Engineering (TKDE)
- IEEE Transactions on Cybernetics
- Information Fusion
- ACM Transactions on Knowledge Discovery from Data (TKDD)
- Pattern Recognition
- Neural Networks
- ACM Transactions on Asian and Low-Resource Language Information Processing
- IEEE Access
- Neural Computation
- Complexity
- Soft Computing
- Journal of Sports Engineering and Technology
- Complex & Intelligent Systems

- Multimedia Tools and Applications
- Big Data

External Conference Reviewer

- AAAI'18-20, ACM CIKM'18-19, Big Data'18, ICDM'18-19, IJCNN'16-17, ITQM'16-17, KDD'18-21, SDM'18-22, TheWebConf (WWW)'20-22, WSDM'20-21

Conference Volunteer

- The Annual Conference of NAACL-HLT, 2021
- Backuping SDM Session Chairs, 2021
- The 35th AAAI Conference on Artificial Intelligence, 2021
- The 26th SIGKDD Conference on Knowledge Discovery and Data Mining, 2020

PATENTS

- Tensorized LSTM with Adaptive Shared Memory for Learning Trends
  W. Cheng, H. Chen, J. Ni, **D. Xu**, and W. Yu
  U.S. Patent. 11,783,181. Oct. 2023
- Spatio Temporal Gated Recurrent Unit
  W. Cheng, H. Chen, and **D. Xu**
  U.S. Patent. 11,461,619. Oct. 2022
- System and Method for Knowledge-Preserving Neural Network Pruning
  E. Yan, **D. Xu**, and Z. Xiao
  U.S. Patent. 11,200,497. Dec. 2021

PATENT APPLICATIONS

- Information-aware Graph Contrastive Learning
  W. Cheng, **D. Xu**, and H. Chen
  U.S. Patent App. 17/728,071. Dec. 2022
- Neural Network Pruning Method and System via Layerwise Analysis
  E. Yan, **D. Xu**, and J. Liu
  U.S. Patent App. 17/107,046. Jun. 2022
- Bank-balanced-sparse Activation Feature Maps for Neural Network Models
  E. Yan, **D. Xu**, and J. Liu
  U.S. Patent App. 17/038,557. Mar. 2022
- Modular Networks with Dynamic Routing for Multi-task Recurrent Modules
  W. Cheng, H. Chen, J. Ni, and **D. Xu**
  U.S. Patent App. 17/158,483. July. 2021
- Unsupervised Multivariate Time Series Trend Detection for Group Behavior Analysis
  W. Cheng, H. Chen, J. Ni, **D. Xu**, and W. Yu
  U.S. Patent App. 16/987,734. Mar. 2021
- Adaptive Neural Networks for Node Classification in Dynamic Networks
  W. Cheng, H. Chen, W. Yu, and **D. Xu**
  U.S. Patent App. 16/872,546. Nov. 2020
- Automated Anomaly Precursor Detection
  W. Cheng, **D. Xu**, H. Chen, and M. Natsumeda
  U.S. Patent App. 16/520,632. Feb. 2020

PROFESSIONAL TALKS

- The Impact of AI on Our Lives and Beyond
  Fo Guang Shan Buddhist Temple, North Carolina, March 2024

- Leveraging Foundation Models for Enhanced Geospatial Analytics and Conservation
  Forest Carbon Solutions Initiative (FCSI), NC State, March 2024
- How LLMs Work and Cutting-Edge Research on Generative AI
  STARS AI Scholars Program, Dec 2023 [remote]
- Sculpting the Future of Collective Growth in Collaborative AI
  Microsoft Research Asia, Beijing, China, Sep 2023 [remote]
- ChatGPT in Corporate Real Estate - Unlocking the Potential [link]
  CoreNet Global, Raleigh, NC, USA, Aug 2023
- Testing Accuracy is Not All You Need: Less Training Cost & More Testing Reliability
  Rutgers University, New Brunswick, USA, Feb 2023
- Resource-efficient Deep Learning: Democratizing AI at Scale
  Pinterest, San Francisco, USA, Aug 2022
- Resource-efficient Deep Learning: Democratizing AI at Scale
  Amazon Search (A9), USA, May 2022
- Resource-efficient Deep Learning: Democratizing AI at Scale
  Vanderbilt University, Nashville, USA, April 2023
- Resource-efficient Deep Learning: Democratizing AI at Scale
  University of Connecticut, Stamford, USA, April 2023
- Parameter Efficiency: Democratizing AI at Scale [Slides]
  Brandeis University, Waltham, USA, Dec 2021
- Chasing Efficiency of Pre-trained Language Models
  Microsoft Research Lab, Redmond, Washington, USA, Jun 2021
- BERT Pruning: Structural vs. Sparse [Slides]
  Brandeis University, Waltham, USA, Apr 2021
- BERT, Compression, and Applications [Slides]
  Xpeng Motors, Mountain View, USA, Apr 2021
- BERT Architecture and Computation Analysis
  Moffett.AI, Los Altos, USA, May 2020.
- Anomaly Precursor Detection via Multi-Instance Contrastive Learning
  NEC Laboratories America, Princeton, USA, May 2019
- Efficient Multiple Instance Learning [Slides]
  NEC Laboratories America, Princeton, USA, May 2018

Honors and Awards

North Carolina State University

- Microsoft Accelerating Foundation Models Research Award 2024
- NCSU Carla Savage Award 2024
- ICCCN Best Paper Award 2023

The Pennsylvania State University

- College of IST Award for Excellence in Teaching Support (top 2) 2019
- Third Place Winner (Eng.) in The 37rd Annual PSU Graduate Exhibition 2022
- NAACL Scholarship 2021
- SIAM Student Travel Award 2021
- KDD Student Registration Award 2020
- AAAI Student Scholarship 2020
- IST Travel Award 2019-2021

University of Chinese Academy of Sciences

- Chinese Academy of Sciences President's Fellowship (the most prestigious award) 2016
- National Graduate Scholarship, China (2% in university) 2016
- Graduate Student Academic Scholarship 2015-2017

Renmin University of China

- First-class Scholarship of Sashixuan Elite Fund, China (5% in university) 2014
- Kwang-hua Scholarship of RUC, China 2014
- Second-class Scholarship of Excellent Student Cadre 2014
- Meritorious Winner in Mathematical Contest in Modeling, USA 2013
- First-class Scholarship of Social Work and Volunteer Service of RUC 2013

EXTRACURRICULAR ACTIVITIES

- IEEE Membership 2023-Present
- ACM Membership 2021-Present
- ACL Membership 2021-Present
- AAAI Student Membership 2019-2021
- Volunteer of Beijing Volunteer Service Federation 2012-2014
- President of Youth Volunteers Association of School of Information 2012-2013
- Leader of National Undergraduate Training Programs 2011-2012