Cheng Guo

c.guo@asu.edu | (+1) (949)-279-8163 | Tempe, AZ
Google Scholar | GitHub

Research Interests

Ph.D. student in Computer Science researching efficient machine learning systems, with a focus on recommender systems, memory-aware model optimization, and sequential recommendation, with peer-reviewed publications at SIGIR, CIKM, and ASP-DAC. My work spans embedding optimization, temporal modeling for personalization, and hardware–software co-design for ML systems. I am interested in scalable learning systems that balance modeling effectiveness with system and hardware constraints.

Education

Ph.D. in Computer Science Aug 2024–Jan 2029 (Expected)
Arizona State University, Tempe, AZ
- Researching efficient machine learning systems especially recommender systems advised by Dr. Jeff Zhang.
- GPA 4.0/4.0.

M.S.E. in Computer Science and Technology Sep 2018–Jun 2021
University of Chinese Academy of Sciences, Beijing, China
- Thesis: Research and Application of Key Technologies for Intelligent Small Commodity Recommendation.
- GPA 3.69/4.0.

B.Eng. in Computer Science and Technology Sep 2014–Jun 2018
Shandong University, Jinan, China
- Thesis: Research on Product Recommendation Algorithms Based on Profiling Technology.
- GPA 88.88/100.

Research Experience

Research Assistant (Ph.D.) Aug 2024–Present
Arizona State University, Tempe, AZ
- (2024–Present) **Embedding Optimization**: Developing and evaluating memory-efficient recommender systems under memory constraints with Dr. Jeff Zhang.
- (2025) **Temporal Embedding**: Designed a time-period-aware embedding regeneration model for session-based recommendation with Dr. Jeff Zhang, accepted by CIKM 2025.
- (2025) **Chiplet-NAS**: Designed the Chiplet-NAS system for efficient AI inference on 2.5D integration with Dr. Yu Cao and Dr. Jeff Zhang, accepted by ASP-DAC 2026.

Student Researcher (Pre-Ph.D.) Sep 2017–Jun 2021
University of Chinese Academy of Sciences (M.S.); Shandong University (B.Eng.)
- (2020) **Session-Based Recommendation (Yiwugo.com Collaboration)**: Conducted session-based recommendation research in collaboration with Yiwugo.com; proposed the hierarchical leaping network accepted at SIGIR 2020 and deployed in production, increasing daily product visits by ∼10%.
- (2021) **Extended Sequential Recommendation Models**: Co-authored extensions accepted at PAKDD 2021 and IJCNN 2021 to improve long-term engagement and personalization.
- (2019) **Explainable Recommender Systems**: Surveyed and synthesized interpretability techniques for recommender systems with Dr. Guandong Xu.
- (2018) **Personalized Collaborative Filtering**: Developed algorithm fusion approaches for real-world recommendation use cases with Dr. Xuemeng Song.
- (2017) **Network Embedding**: Explored representation learning methods for graph-structured data with Dr. Zhumin Chen.

Publications

- Cheng Guo, Pragnya Sudershan Nalla, Nikhil Kumar Cherukuri, Rui Xue, Sachin S. Sapatnekar, Chaitali Chakrabarti, Yu Cao, and Jeff Zhang. "Chiplet-NAS: Chiplet-Aware Neural Architecture Search for Efficient AI Inference on 2.5D Integration". In *ASP-DAC 2026*.
- Cheng Guo, Rui Xue, and Jeff Zhang. "Time-Period-Aware Embedding Regeneration for Session-Based Recommendation". In *CIKM 2025*.
- Rui Xue, Wenming Li, Haibin Wu, Cheng Guo, Yi Li, Xiaochun Ye, and Dongrui Fan. "A High-Performance Dataflow-Based ORB Extractor Accelerator for SLAM". In *ISLPED 2025*.
- Asmer Hamid Ali, Amitesh Sridharan, Cheng Guo, William Hwang, Wilman Tsai, Jeff Zhang, Yiran Chen, Shan X. Wang, and Deliang Fan. "FP-SMR: A Fully Digital Floating-Point Processing-in-SAS-MRAM for Session-Based Recommender System". In *GLSVLSI 2025*.
- Mengfei Zhang, Cheng Guo, Jiaqi Jin, Mao Pan, and Jinyun Fang. "Modeling Hierarchical Intents and Selective Current Interest for Session-Based Recommendation". In *PAKDD 2021*.
- Mengfei Zhang, Cheng Guo, Jiaqi Jin, Mao Pan, and Jinyun Fang. "Sequential Recommendation with Context-Aware Collaborative Graph Attention Networks". In *IJCNN 2021*.
- Cheng Guo, Mengfei Zhang, Jinyun Fang, Jiaqi Jin, and Mao Pan. "Session-Based Recommendation with Hierarchical Leaping Networks". In *SIGIR 2020*.

Industry Experience

Microsoft - Software Engineer Jul 2021–Dec 2022

Beijing, China

- Led end-to-end design and launch of the Bing Donation NPO globalization initiative, coordinating cross-functional teams for worldwide rollout.
- Delivered Bing Guest Donation support enabling contributions from users without Microsoft accounts.
- Prototyped the Edge Wallet express checkout feature (edge://wallet) on the Chromium-based Edge browser to streamline user-facing e-commerce flows.

Technical Strengths

Core Skills: Python, PyTorch, Git, Docker

Systems & Optimization: C/C++, Optuna, NVIDIA Triton Inference Server

Prior Research & Industry Experience: Faiss, Spark, Kafka, Java, SQL, Redis, Elasticsearch, HBase

Awards & Honors

DAC Young Fellow (Design Automation Conference)	2025
Outstanding Student, UCAS	2020
Outstanding Scholarship, ICT-CAS	2019
American Mathematical Contest in Modeling (Honorable Mention)	2017
Outstanding Student Scholarship, Shandong University	2016
National Inspirational Scholarship (Chinese Ministry of Education)	2015
National College Student Mathematics Competition (Provincial First Prize)	2015