XUYING NING

(+1) 447-902-6907 | xuyingn2@illinois.edu | 614 E Daniel St, Champaign, IL 61820

EDUCATION

University of Illinois Urbana-Champaign Illinois, USA
M.S. in Computer Science (Continuing to Ph.D. Track) Aug 2024 – May 2026

- GPA: 4.0/4.0
- Advisors: Prof. Jingrui He
- Recognitions: Siebel Scholar (Highest Honer)

Xi'an Jiaotong University Xi'an, China
B.Eng. (Hons) in Artificial Intelligence Sep 2020 – Jul 2024

- GPA: 3.91/4.3 (Rank Top 5%)
- Advisor: Prof. Nanning Zheng
- Honors Youth Program of Qian Xuesen Honors College (Enrolled in University at the age of 15)
- Artificial Intelligence (AI) Experimental Class and the Elite Program
- Recognitions: Outstanding Student Award, MEGVII Scholarship (top 3% of students in Artificial Intelligence)

RESEARCH INTEREST

- My current research interests lie at the intersection of (multimodal) large language models and graph foundation models, with applications in recommendation systems and agentic search

PUBLICATIONS

- **Xuying Ning**, Dongqi Fu, Tianxin Wei, Mengting Ai, Jiaru Zou, Ting-Wei Li, Hanghang Tong, Yada Zhu, Hendrik Hamann, Jingrui He, ***"****MC-Search: Evaluating and Enhancing Multimodal Agentic Search with Structured Long Reasoning Chains****",*** **ICLR 2026 Oral**
- **Xuying Ning**, Dongqi Fu, Tianxin Wei, Wujiang Xu, Jingrui He, "*Graph4MM: Weaving Multimodal Learning with Structural Information*", **ICML 2025**
- **Xuying Ning**, Wujiang Xu, Tianxin Wei, Xiaolei Liu, "*i^2VAE: Interest Information Augmentation with Variational Regularizers for Cross-Domain Sequential Recommendation*", **UAI 2025**
- **Xuying Ning**, Dongqi Fu, Tianxin Wei, Mengting Ai, Jiaru Zou, Ting-Wei Li, Jingrui He, "*Search-MM: Benchmarking Multimodal Agentic RAG with Structured Reasoning Chains*", Accepted by **NeurIPS 2025 LLM Evaluation Workshop**
- Junyu Zhang, Runpei Dong, Han Wang, **Xuying Ning**, Haoran Geng, Peihao Li, Xialin He, Yutong Bai, Jitendra Malik, Saurabh Gupta, Huan Zhang, "*AlphaOne: Reasoning Models Thinking Slow and Fast at Test Time*", **EMNLP 2025 Main**
- Wujiang Xu, Yunxiao Shi, Zujie Liang, **Xuying Ning**, Kai Mei, Kun Wang, Xi Zhu, Min Xu, Yongfeng Zhang, "*iAgent: LLM Agent as a Shield between User and Recommender Systems*", **ACL Findings 2025**
- Wujiang Xu, Qitian Wu, Zujie Liang, Jiaojiao Han**, Xuying Ning**, Wenfang Lin, Linxun Chen, Feng Wei, Yongfeng Zhang, "*SLMRec: Empowering Small Language Models for Sequential Recommendation*", **ICLR 2025**
- Wujiang Xu, **Xuying Ning**, Wenfang Lin, Mingming Ha, Qiongxu Ma, Linxun Chen, Bing Han, Minnan Luo, "*Towards Open-World Cross-Domain Sequential Recommendation: A Model-Agnostic Contrastive Denoising Approach*", **ECML-PKDD 2024**
- Tianxin Wei*, **Xuying Ning***, Xuxing Chen, Yan Xie, Ruizhong Qiu, Zhigang Hua, Jingrui He, "*CoFiRec: Structuring Token Granularity for Generative Recommendation*", **WWW 2026** (under review)
- Youru Li, **Xuying Ning**, Zhenfeng Zhu, Linxun Chen, Hanqiu Wang, Minnan Luo, Bo Zheng, Qinghua Zheng, Yao Zhao, "*Towards Improving Trustworthiness of Personalized Online Service: A Knowledge-augmented Method*", **TKDE** (under review)
- Xinrui He, Ting-Wei Li, Tianxin Wei, **Xuying Ning**, Xinyu He, Wenxuan Bao, Hanghang Tong, Jingrui He, "*FeDecider: An LLM-Based Framework for Federated Cross-Domain Recommendation*", **KDD 2026** Research Track August (under review)
- Tianxin Wei, Zhichen Zeng, Ruizhong Qiu, Zhining Liu, **Xuying Ning**, Xinrui He, Wenxuan Bao, Qi He, Xianfeng Tang, Hanghang Tong, Jingrui He, "Inference Scaling of LLM Ensembling: Bridging Token Spaces with Token Translation", **ICLR 2026** (under review)

PROFESSIONAL EXPERIENCE

Ant Group Hangzhou, China
Machine Learning Engineer Intern Jul 2023 – Apr 2024

- Analyzed MYbank's financial product recommendation systems; developed a novel De-biasing Knowledge Graph-based Recommendation System (DiKGRS) framework, which effectively addresses exposure bias in existing recommendation algorithms, using Knowledge Graph-guided data augmentation and a deep value network

- Created a model-agonistic attention-based contrastive denoising framework (MACD) for cross-domain sequential recommendation, aiming to tackle the performance challenges associated with cold-start and long-tail users in the company's large-scale recommendation system
- DiKGRS is deployed on MYbank's fund recommendation platform; in a fourteen-day online A/B test, the MACD framework exhibited great enhancements in a large-scale loan benefit recommendation scenario, resulting in a 10.29% increase in average exposure and a 6.28% boost in conversion rate

Ruitian Capital Shanghai, China
Quantitative Research Intern – ML Team Feb 2023 – May 2023

- Individually constructed an ML pipeline predicting stock yields for a five-day holding period in the Chinese stock market, which included data preprocessing, feature engineering, purged time series data cross-validation, model design, training, ensemble, and inference; gained experience in multiprocessing and distributed training
- Developed stock indiscriminate regression models, LightGBM and AutoEncoder+MLP; implemented a sequential model Bi-LSTM+Multi-head Attention module, employing a window of five consecutive days of each stock's historical factor data; applied Optuna for Bayesian hyperparameter optimization and neural architecture searches, the ensemble results outperformed the baseline strategy on both MSE (by 33%) and Information Coefficient

RESEARCH EXPERIENCE

University of Illinois Urbana-Champaign Illinois, USA
Siebel School of Computing and Data Science Aug 2024 – Now
Advised by Prof. Jingrui He

Graph4MM: Weaving Multimodal Learning with Structural Information

- Proposed a novel graph-based multimodal learning framework that injects multi-hop structural priors into foundation models for fine-grained multimodal reasoning
- Designed Hop-Diffused Attention, which integrates multi-hop neighbor information into self-attention through hop-based diffusion and causal masking
- Introduced MM-QFormer, a multi-mapping querying transformer for structured cross-modal fusion of image and text guided by graph context
- Achieved significant performance gains (avg. +6.93%) over strong baselines including BLIP-2 and large VLMs across both generative and discriminative tasks on four real-world multimodal datasets
- First-author full research paper accepted to the Forty-Second International Conference on Machine Learning (ICML 2025)

CoFiRec: Coarse-to-Fine Autoregressive Recommendation via Structured Tokenization

- Proposed CoFiRec, a novel generative recommendation framework that aligns structured item semantics with autoregressive preference modeling through coarse-to-fine tokenization
- Designed the CoFiRec Tokenizer using level-wise vector quantization over item metadata and collaborative signals, enabling interpretable and expressive semantic representations
- Reformulated autoregressive decoding to follow a hierarchical generation order, improving alignment between user preferences and token-level outputs
- Conducted theoretical analysis and empirical studies demonstrating that coarse-to-fine generation reduces semantic dissimilarity and improves accuracy, achieving +7.27% performance gain on multiple public datasets
- Co-first-authored submission currently under review at the Web Conference 2026

Ant Group Hangzhou, China
MYbank Intelligent Engine Department Jul 2023 – Jul 2024
Research-oriented Machine Learning Engineer Intern

Cross-Domain Sequential Recommendation: A Model-Agnostic Contrastive Denoising Approach

Co-advised by Prof. Nanning Zheng and Prof. Minnan Luo from Xi'an Jiaotong University

- Proposed a model-agnostic framework leveraging user's auxiliary behaviors to enhance cross-domain sequential models in an open-world setting
- Collaborated in designing an attention-based Denoising Interest-aware Network with a Contrastive Information Regularizer to remove noise from auxiliary behaviors and capture evolving user interest across domains
- Developed denoising and fusion gate units for integrating inner- and cross-domain interest representations.
- Conducted extensive experiments on Amazon datasets, achieving significant improvements ranging from 3.19% to 11.11% compared to the previous state-of-the-art baseline. In a fourteen-day online A/B test, our method demonstrated a 10.29% increase in average exposure, a 6.28% increase in conversion rate, and a 1.45% increase in CVR across three domains
- Co-authored research paper published at ECML-PKDD 2024; ArXiv preprint is available

i^2VAE: Interest Information Augmentation with Variational Regularizers for Cross-Domain Sequential

- Proposed i²VAE, a mutual-information-regularized variational autoencoder that disentangles and aligns intra- and cross-domain user interests in partially overlapping user scenarios
- Introduced an interest augmentation module that generates denoised pseudo-sequences to improve robustness and long-tail generalization in sequential recommendation

- Achieved significant performance gains over state-of-the-art SDSR, CDR, and CDSR models, especially for cold-start and long-tailed users on large-scale Amazon datasets
- First-author full paper accepted to the Conference on Uncertainty in Artificial Intelligence (UAI 2025); preprint available on arXiv

Trustworthy Recommendation System: A Knowledge Graph-Driven Data Augmentation Strategy
Co-advised by Prof. Qinghua Zheng and Prof. Minnan Luo from Xi'an Jiaotong University

- Proposed a novel de-biasing data augmentation framework, DiKGRS, leveraging facts in item-to-item knowledge graphs (KGs) as unbiased prior to measuring item similarities based on item-centering multi-hop subgraph's meta-path and semantics representation
- Integrated user profiles into a DeepFM-like Deep Value Network, which learns sample weights within the BPR loss function to enhance recommendation fairness for minority user groups
- Independently developed the project and conducted extensive experiments and ablation studies on MovieLens1M, LastFM, and MyBank's industrial datasets, resulting in noticeable improvements in both recommendation accuracy (by 2.42%) and exposure fairness metrics (by 54.29%)
- The co-authored paper is under review at IEEE Transactions on Pattern Analysis and Machine Intelligence

SLMRec: Empowering Small Language Models for Sequential Recommendation
Co-advised by Prof. Yongfeng Zhang from Rutgers University

- Challenged the necessity of large language models (LLMs) in sequential recommendation by evaluating the scaling law's applicability, identifying potential over-parameterization
- Applied vanilla knowledge distillation techniques to efficiently transfer knowledge, achieving robust performance with models under 1 billion parameters compared to larger models
- Demonstrated significant performance gains with smaller models, achieving competitive results and up to 6.6x/8.0x speed improvements in training and inference times
- Co-authored paper published at ICLR 2025; also available on arXiv

Columbia University New York, USA
Computer Science Department Sep 2022 – Dec 2022
Research Project Advised by Prof. Peter Belhumeur
Masked Face Dataset Generation and Masked Face Recognition

- Collaborated to solve the problem of small sample face recognition in the case of occlusion; coordinated the project design, including Dataset EDA, data augmentation, fine-tuning of pre-trained Inception-ResNet v1(VGGFace), and ResNet 50 (ImageNet) on the mask-simulated face dataset; and proposed the "Exponential Margin" approach to determine whether an incoming new image belongs to our recognized targets
- Research results: Achieved a recognition accuracy of 93% (potentially still underfitting) with fine-tuned Inception-ResNet-v1 model and the best accuracy of 95% with the ResNet-50 model

National University of Singapore Singapore
School of Computing Summer Workshop 2022 May 2022 – Jul 2022
Member of AI/ML in Financial Services Summer Workshop
TradeGuard No.5, a Comprehensive DL-Driven Stock Evaluation and Portfolio Recommendation Platform

- Developed and utilized deep learning/machine learning models to predict future U.S. stock closing prices; created an LSTM with a News Sentiment Prediction Model based on Random Forrest Classifier and TF-IDF for prediction of the next day's stock returns, achieving an impressive 95% accuracy on a dataset of 26 U.S. stocks
- Innovated and developed a new time stamp encoding mechanism and a parametrical embedding layer in Transformer to forecast stock prices for a 30-day period, reaching an accuracy of 87% on the same dataset
- Research results: Recognized with the Best Project prize for the comprehensive deep learning-driven stock prediction techniques and received an A+ rating in personal performance evaluation

AWARDS & HONORS

- Siebel Scholar (Class of 2026) 2025
- Xi'an Jiaotong University Outstanding Student Award 2022
- Best Project Prize of National University of Singapore SOC Summer Workshop 2022
- Xi'an Jiaotong University Undergraduate Academic Scholarship 2022
- MEGVII Scholarship (less than 3% of students in Artificial Intelligence) 2021
- Excellent Student Cadre Award of Xi'an Jiaotong University 2021
- Second Prize of National Undergraduate Mathematical Contest in Modeling (Shaanxi Division) 2021
- Honorable Mention of Interdisciplinary Contest in Modeling (ICM) 2021

ADDITIONAL INFORMATION

- Technical Skills: Proficient in Python (including PyTorch and TensorFlow), C++, SQL
- Languages: Native in Mandarin, Proficient in English (TOEFL 115, Speaking 27)