

## Education & Employment

- 2022.07 – present    📌 Research Engineer, **Alibaba Group** on Recommender Systems.
- Research Advisor: Chao Du (now at Sea AI).
  - Foster Business Growth by 40% in revenue. Establish tech impact (KDD23, Press Coverage).
- 2019.09 – 2022.07    📌 M.Sc. Computer Science, **ShanghaiTech University**.
- Supervisor: Xuming He, Associate Professor.
  - GPA: 3.83/4.0 (Top 2%).
  - Honors: **Outstanding Graduates** of Shanghai. (No. 20221442303013, 1% province-wide)
- 2015.09 – 2019.07    📌 B.Sc. Computer Science (Information Security), **Wuhan University**.
- GPA: 3.88/4.0 (Top 1%=2/150)
  - Honors: **National Scholarship** (in 2017 No. 11959, 0.2% nation-wide)

## Research Experience

I have research experience in various topics, including RL, Robust Generalization, Transformers, Learning to Rank.

### • Rapid and Robust Generalization of RL

- KDD 2023    📌 **Adversarial Constrained Bidding via Minimax Regret Optimization.** [paper]  
First Author    Summary: ERM-based policies fail in black-box adversarial bidding environments due to train-test mismatch. To address this, we proposed to train toward maximizing the worst-case performance, resulting in a Minimax Optimization problem. However, the minimax game is non-trivial to solve due to black-box environments. To address this, we further proposed to reconstruct world model and utilize Variational Bayes to embed the unknown environment.
- KDD 2022    📌 **ROI-Constrained Bidding via Curriculum-Guided Bayesian Reinforcement Learning.** [paper]  
First Author    Summary: We study how to enforce constraints for policies in non-stationary bidding environments. Our results show that Lagrangian Duality methods cannot ensure constraints due to the un-controlled soft penalty on constraint violations. Instead, we utilize the indicator function that can sets hard barrier for feasibility. To alleviate inefficient policy learning due to sparse reward, we propose a curriculum learning approach that guides policy search by scheduling relaxed optimization problems. To handle non-stationarity, we propose to model uncertainty through Bayesian Inference.
- AAMAS 2020    📌 **Context-aware Task Reasoning for Efficient Meta-Reinforcement Learning.** [paper]  
First Author    Summary: We point out an under-explored aspect of Meta-Reinforcement Learning (Meta-RL) : active task-exploration is crucial to achieve rapid task learning. Hence we propose to decompose Meta-RL into task-exploration, task-inference and task-execution. To jointly learn these components, we propose a unified Variational EM Learning Framework. The resulting method can efficiently explores a novel task and effectively reasons about the task context for rapid adaptation.

### • Vision and Language

- ICCV 2023    📌 **Grounded Image-Text Matching with Mismatched Relation Reasoning.** [paper]  
Third Author    Summary: We find that pre-trained vision-language transformers are limited in relation reasoning, length generalization and data efficiency. Therefore, We propose a neuro-symbolic approach that performs belief propagation over the language structure cues for reasoning. Our results show that the length extrapolation issue is mitigated using graph entity-level attention (v.s. token-level attention) due to less train-test mismatch.

### • Human Preference Learning

- Work    📌 Preference Estimation and Ranking from Discrete and Continuous Feedback.  
Summary: The goal is to estimate and rank user-item affinities (i.e., preference) according to multiple objectives, including discrete feedback, e.g., clicks, and continuous feedback, e.g., GMV, income. However, such feedback can be noisy, e.g., robot clicks, and can have high variance, e.g., GMV. Key techniques include: Pair-wise Loss, Ordinal Regression.

## Misc

- Services    📌 Peer review experience in KDD 2022, KDD 2023.
- English    📌 Fluent Writing. Scores: CET4 661/710, CET6 611/710, TOFEL 106/120, GRE 326/340.
- Coding    📌 Proficiency in PyTorch, TensorFlow and common libraries for ML research.

## Research Interests

---

My current research focus is on Generative AI. Recently I am interested and have surveyed the following directions of LLMs:

- *Robust & Trustworthy Generation.*
  - Retrieval-Augmented Generation (RAG):
    - \* handling retrieval errors;
    - \* effectively using long context (for factual consistency), e.g., addressing ‘Lost in the Middle’ phenomenon.
  - Factuality & Hallucinations , e.g.,
    - \* grounded evidence, citations;
    - \* handling free-form document (texts and tables)
    - \* long-form output, e.g., Long-Form QA.
  - Finetuning & Alignment
    - \* handling biases, e.g., confirmation bias;
    - \* handling alignment black-box, inconsistent or conflicting judgement
    - \* second-order oversight, multi-agent alignment
- *Knowledge Editing.* When pre-trained LLM contain misinformation, how to inject new facts with entailments?
- *LLM Lifelong Learning.*