Ye Yuan

ye.yuan3@mail.mcgill.ca | (438) 351-1806 | GitHub | LinkedIn | Google Scholar

EDUCATION

- **McGill University & Mila – Quebec AI Institute** Montreal, Quebec, Canada
  *Doctor of Philosophy (PhD) in **Computer Science*** | CGPA: 3.89/4.00 January 2023 – Present
  Supervisor: Professor **Xue (Steve) Liu** (with Professor Adriana Romero Soriano as my supervision committee member)
  Awards: DAAD AINeT Fellowship | Bank of Montreal (BMO) Responsible AI Senior Scholar | BMO Responsible AI Fellowship | ICLR 2025 Financial Assistance | NeurIPS 2023 Scholar Award | McGill Faculty of Science Graduate Scholarship | McGill Graduate Excellence Awards
- **McGill University** Montreal, Quebec, Canada
  *Bachelor of Science (BSc) in **Honours Computer Science*** | CGPA: 3.95/4.00 September 2019 - December 2022
  Awards: First Class Honours | Tomlinson Undergraduate Award | Faculty of Science Scholarship | Dean's Honour List

PUBLICATIONS / PREPRINTS

- **Importance-Aware Co-Teaching for Offline Model-Based Optimization**
  Paper Link | GitHub Repository | Blog Promoted by Mila
  Published Venue: The Thirty-Seventh Annual Conference on Neural Information Processing Systems (NeurIPS 2023)
  *Authors: **Ye Yuan***, Can Chen*, Zixuan Liu, Willie Neiswanger, Xue Liu*
- **Learning to Extract Structured Entities Using Language Models**
  Paper Link | GitHub Repository | Official Recording of the Oral Presentation
  Published Venue: The 2024 Conference on Empirical Methods in Natural Language Processing (EMNLP 2024)
  *Authors: **Ye Yuan***, Haolun Wu*, Liana Mikaelyan, Alexander Meulemans, Xue Liu, James Hensman, Bhaskar Mitra*
  - Selected as an **oral presentation** at the **main conference**, 168 out of 2271 accepted papers (7%)
- **ParetoFlow: Guided Flows in Multi-Objective Optimization**
  Paper Link | GitHub Repository | X Post Promoted by Mila
  Published Venue: The Thirteenth International Conference on Learning Representations (ICLR 2025)
  *Authors: **Ye Yuan***, Can Chen*, Christopher Pal, Xue Liu*
- **Design Editing for Offline Model-based Optimization**
  Paper Link | GitHub Repository
  Published Venue: Transactions on Machine Learning Research (TMLR 2025)
  *Authors: **Ye Yuan***, Youyuan Zhang*, Can Chen, Haolun Wu, Zixuan Li, Jianmo Li, James J. Clark, Xue Liu*
- **Understanding 6G through Language Models: A Case Study on LLM-aided Structured Entity Extraction in Telecom Domain**
  Paper Link
  *Authors: **Ye Yuan**, Haolun Wu, Hao Zhou, Xue Liu, Hao Chen, Yan Xin, Jianzhong (Charlie) Zhang*
- **Offline Model-Based Optimization: Comprehensive Review**
  Paper Link | GitHub Repository | X Post Promoted by Mila
  Authors: Minsu Kim, Jiayao Gu*, **Ye Yuan**, Taeyoung Yun, Zixuan Liu, Yoshua Bengio, Can Chen*
- **Large Language Model (LLM) for Telecommunications: A Comprehensive Survey on Principles, Key Techniques, and Opportunities**
  Paper Link | LinkedIn Post Promoted by IEEE COMST
  Published Venue: IEEE Communications Surveys & Tutorials (IEEE COMST 2024)
  *Authors: Hao Zhou, Chengming Hu, **Ye Yuan**, Yufei Cui, Yili Jin, Can Chen, Haolun Wu, Dun Yuan, Li Jiang, Di Wu, Xue Liu, Charlie Zhang, Xianbin Wang, Jiangchuan Liu*
- **Generative AI as a Service in 6G Edge-Cloud: Generation Task Offloading by In-context Learning**
  Paper Link
  Published Venue: IEEE Wireless Communications Letters (IEEE WCL 2024)
  *Authors: Hao Zhou, Chengming Hu, Dun Yuan, **Ye Yuan**, Di Wu, Xue Liu, Zhu Han, Jianzhong Zhang*
- **Large Language Models for Wireless Networks: An Overview from the Prompt Engineering Perspective**
  Paper Link
  Published Venue: IEEE Wireless Communications (IEEE WCM 2024)
  *Authors: Hao Zhou, Chengming Hu, Dun Yuan, **Ye Yuan**, Di Wu, Xi Chen, Hina Tabassum, Xue Liu*
- **Large Language Model Enabled In-context Learning for Wireless Network Optimization: A Case Study of Power Control**

Paper Link
*Authors: Hao Zhou, Chengming Hu, Dun Yuan, **Ye Yuan**, Di Wu, Xue Liu, Charlie Zhang*

- **Retrieval-Augmented Generation for Natural Language Processing: A Survey**
Paper Link
*Authors: Shangyu Wu, Ying Xiong, Yufei Cui, Haolun Wu, Can Chen, **Ye Yuan**, Lianming Huang, Xue Liu, Tei-Wei Kuo, Nan Guan, Chun Jason Xue*

WORK EXPERIENCES

- **RBC (Royal Bank of Canada) Borealis** **May 2025 - Present**
Machine Learning Researcher Intern Toronto, Ontario, Canada
  - Worked for the embedding base context aware reranker project. By integrating rerankers, RAG systems can strike a better balance between speed and precision, leading to more reliable and efficient outputs in various applications.
  - Investigated the efficiency and efficacy of rerankers and developed new reranking algorithm based on embeddings.

- **Samsung Research America** **January 2025 - Present**
Research Assistant for Project of Large Language Model-Driven Efficient and Flexible Network Management for NextG Cellular Communications Remote
(Supported by Samsung Global Research Outreach program)
  - Applied the entity extraction strategies in automatically constructing knowledge bases for telecommunications.
  - Benchmarked different models with delicately configured simulations to test their real-world performance.

- **Microsoft Research** **March 2023 - Present**
Research Assistant for Project Alexandria (Supported by MSR-Mila Research Grant) Montreal, Quebec, Canada
  - Introduced and formalized the task of structured entity extraction within the realm of strict information extraction.
  - Proposed an evaluation metric called Approximate Entity Set OverlaP (AESOP) with numerous variants tailored for assessing structured entity extraction and aligned more with human preferences.
  - Created a multi-stage decoding strategy for structured entity extraction, improving the effectiveness and efficiency.
  - Led the study of the properties of generative models and discriminative models for entity linking tasks.

- **Huawei Noah's Ark Lab Canada** **April 2023 – April 2025**
Associate Researcher Intern at Reasoning-Decision-Making Team and NLP Team Montreal, Quebec, Canada
  - ***Recipient of Outstanding Contribution Award in R&D Peripheral Fields | Overseas Business Contribution Award | Canada Research Institute President's Spot Award***
  - Adapted the soft prompting to distribution shift problems and empowered a single model for multiple scenarios.
  - Explored novel model compression ideas like LoRA-pruning and Mixture of Depth model to save calculations.
  - Took strategy insights into cutting-edge LLM methods and core techniques from leading unicorn startups.

- **Mila – Quebec AI Institute** **May 2024 – December 2024; May 2025 – Present**
Mentor for Professional MSc Student's Internship Montreal, Quebec, Canada
  - Supervised professional Master of Science students for their internships' projects at Deep River and Bell Canada.
  - Provided guidance and suggestions for entity-linking, fine-tuning, knowledge distillation, and RAG for LLMs.
  - Offered supervision and advice for document extraction, named entity extraction, and coreference resolution.

ADDITIONAL

- **Reviewer Services**
  - **Conferences:** NeurIPS 2025; ACL Rolling Review; ICML 2025; ICLR 2025; AISTATS 2025; NeurIPS 2024.
  - **Workshops:** ICLR 2025 DeLTa Workshop; AAAI 2025 Undergrad Consortium; AAAI 2025 KnowFM Workshop; NeurIPS 2024 SafeGenAi Workshop; ICML 2024 FM-Wild Workshop.
  - **Journals:** Transactions on Machine Learning Research (TMLR); ACM Transactions on Information Systems.
- **Tutorships**
  - Teaching Assistants of Foundations of Programming and Introduction to Computer Science for four semesters.
  - Mentor and Judge for McGill McHacks Hackthon 9, 10, 11.
- **Relevant Courses**
  - Taken during pursuing PhD: Computer Networks, Matrix Computations, Computational Biology Methods and Research, Machine Learning for Biomedical Data, Natural Language Understanding with Deep Learning, Deep Learning, Introduction to Ethics of Intelligent System, Probabilistic Graphical Models.
  - Taken during pursuing BSc: Cryptography and Data Security, Intelligent Software System, Applied Machine Learning, Reinforcement Learning, Operating Systems, Natural Language Processing, Software Design, Modern Computer Games, Data Structures, Probability, Numerical Analysis, Linear Algebra, Discrete Mathematics, Calculus.