

Haolun Wu

Ph.D. in Computer Science, McGill University, Mila- Quebec AI Institute

Visiting Scholar, Computer Science, Stanford University

Canada Governor General's Gold Medal Recipient

FRQNT Ph.D. Fellowship, Borealis AI Fellowship

Room - 305, McConnell Engineering Building

School of Computer Science, McGill University

☎ (+1) 514-431-5472

✉ haolunwu@cs.stanford.edu; haolun.wu@mail.mcgill.ca

📄 <http://haolun-wu.github.io/>

Last updated in 8/2026

Research Interests

Areas: Personalization, Information Retrieval, Trustworthy AI, AI Alignment

My research focuses on trustworthy and personalized AI that learns from human feedback to better understand, represent, and serve human needs. I work across both *micro-level* dimensions (e.g., data quality, values, personalization, and user modeling) and *macro-level* objectives (e.g., responsibility, trustworthiness, social norms, and interdisciplinary impact). My recent interests include **mechanistic interpretability, modeling long-horizon human-human/AI interaction, and safe evolution of AI agents**. My research has been supported by multiple fellowships and grants, and published in top venues including NeurIPS, ICLR, ICML, WWW, SIGIR, KDD, EMNLP, TMLR, TKDE, CHI, and CSCW. During my Ph.D., I interned and collaborated closely with Microsoft Research, Samsung AI Research America, Google Research, and Google DeepMind. I am a recipient of Canada Governor General's Gold Medal.

Academic Training

2021-2026 **Mila & McGill University, Montreal, Canada.**

Ph.D., School of Computer Science.

Thesis: *Human-centric Information Access through Individual and Population Level Alignment*

Supervisory committee: [Steve Liu](#), [Fernando Diaz](#), [Laurent Charlin](#), [Joëlle Pineau](#)

To the top graduate across the entire university: [Governor General's Gold Medal](#) & D.W. Ambridge Prize.

2024-2025 **Stanford University, [STAIR Lab](#), Palo Alto, USA.**

Visiting Scholar, Department of Computer Science.

Supervisor: [Sanmi Koyejo](#)

2019-2021 **Mila & McGill University, Montreal, Canada.**

M.Sc., School of Computer Science.

Thesis: *Balancing Fairness in Multi-stakeholder Recommendation via Multi-objective Optimization*

Published at [ACM Transactions on Information Systems \(TOIS\)](#).

2018 **University of California, San Diego (UCSD), La Jolla, USA.**

Undergraduate Exchange.

2015-2019 **Northeastern University, Shenyang, China.**

B.E. Computer Science and Engineering.

Outstanding Graduation Thesis (top 1%). Outstanding Student for all years.

Relevant Industrial Experience

Jan. 2025- **Samsung Research America, Montreal, Canada.**

Dec. 2025 **Research Assistant.**

Project collaborator: [Hao Chen](#), [Hao Zhou](#).

- Research collaboration between McGill University and Samsung Research on applying large language models and multi-agent systems to next-generation telecommunications.

Feb. 2023- **Microsoft Research**, *Cambridge, UK*.

Aug. 2024 **Student Researcher**, [Alexandria](#) Team.

Manager: [John Winn](#). Project mentor: [Bhaskar Mitra](#).

- Propose a structured entity extraction framework that reformulates information extraction as an entity-centric task, introduces the AESOP metric for entity-level evaluation, and presents the multi-stage MuSEE model that leverages LLMs to achieve more effective and efficient extraction.

Aug. 2023- **Google Research / Google DeepMind**, *Mountain View, USA*.

Feb. 2024 **Student Researcher**.

Project mentor: [Cicero Nogueira dos Santos](#).

- Work on knowledge localization in LLMs for both dense architectures and sparse architectures (i.e., Mixture-of-Experts).

Nov. 2022- **Google Research**, *Mountain View, USA*.

Feb. 2023 **Research Intern**.

Manager / mentor: [Craig Boutilier](#).

- Propose density-based user representation (DUR) through Gaussian Process Regression, for better personalized multi-interest retrieval.

Jul. 2022- **Bell Canada**, *Montreal, Canada*.

Jun. 2023 **Research Intern**.

Team: Dynamic Call Routing Team

- Propose a fairness-aware learning-to-rank framework for fair agent-customer matching in call center to balance agent workloads and improve global customer value.

Sep. 2020- **Microsoft Research**, *Montreal, Canada*.

Jun. 2021 **Student Researcher**, **FATE** (*Fairness, Accountability, Transparency, and Ethics*) Team.

Manager / mentor: [Fernando Diaz](#).

- Propose a multi-objective optimization framework for fairness-aware recommendation, Multi-FR, that adaptively balances accuracy and fairness for various stakeholders with Pareto optimality guarantee.

Publications - Full list on my [Google Scholar](#) profile.

During my Ph.D., I have published 45+ peer-reviewed works, including 10+ papers as (co-)first/corresponding author. These top-tier venues include but not limited to NeurIPS, ICML, ICLR, ACL, EMNLP, SIGIR, KDD, WWW, CHI, CSCW, TKDE, EMLR, etc. My work has received 2,200+ citations with an h-index of 20 on Google Scholar by 8/2026.

arXiv Zhuohan Xie, Yuyang Dai, ..., **Haolun Wu**, ..., Xue Liu, Preslav Nakov. “Overview of FinMMEval 2026 Task 1: Multilingual Financial Multiple-Choice Question Answering”. arXiv preprint arXiv:2607.19856.

arXiv Zhuohan Xie, Xueqing Peng, ..., **Haolun Wu**, ..., Xue Liu, Preslav Nakov. “Overview of FinMMEval 2026 Task 2: Multilingual Financial Short-Answer Question Answering”. arXiv preprint arXiv:2607.19867.

arXiv Ye Yuan, Weien Li, ..., **Haolun Wu**, ..., Philip S. Yu, Xue Liu. “Discrete Diffusion Models: A Unified Framework from Tokenization to Generation”. arXiv preprint arXiv:2607.13431.

- arXiv Weien Li, Rui Song, Zeyu Li, Haochen Liu, Gonghao Zhang, Difan Jiao, Zhenwei Tang, Bowei He, **Haolun Wu**, Xue Liu, Ye Yuan. “*MINER: Mining Multimodal Internal Representation for Efficient Retrieval*”. arXiv preprint arXiv:2605.06460.
- arXiv Dun Yuan, Fuyuan Lyu, Ye Yuan, Weixu Zhang, Bowei He, Jiayi Geng, Linfeng Du, Zipeng Sun, Yankai Chen, Changjiang Han, Jikun Kang, Xi Chen, **Haolun Wu**, Xue Liu. “*Beyond Message Passing: A Semantic View of Agent Communication Protocols*”. arXiv preprint arXiv:2604.02369.
- arXiv **Haolun Wu***, Zhenkun Li*, Lingyao Li*. “*Can LLM Agents Really Debate? A Controlled Study of Multi-Agent Debate in Logical Reasoning*”. arXiv preprint arXiv:2511.07784.
- arXiv Lingyao Li*, **Haolun Wu***, Zhenkun Li*, Jiabei Hu, Yu Wang, Xiaoshan Huang, Wenyue Hua, Wenqian Wang. “*PartnerMAS: An LLM Hierarchical Multi-Agent Framework for Business Partner Selection on High-Dimensional Features*”. arXiv preprint arXiv:2509.24046.
- arXiv Qiyuan Zhang, Fuyuan Lyu, Zexu Sun, Lei Wang, Weixu Zhang, Wenyue Hua, **Haolun Wu**, Zhihan Guo, Yufei Wang, Niklas Muennighoff, Irwin King, Xue Liu, Chen Ma. “*A Survey on Test-Time Scaling in Large Language Models: What, How, Where, and How Well*”. arXiv preprint arXiv:2503.24235.
- arXiv Fuyuan Lyu, Xing Tang, Dugang Liu, **Haolun Wu**, Chen Ma, Xiuqiang He, and Xue Liu. “*Feature Representation Learning for Click-through Rate Prediction: A Review and New Perspectives*”. arXiv preprint arXiv:2302.02241.
- arXiv Chengming Hu, Xuan Li, Dan Liu, **Haolun Wu**, Ju Wang, and Xi Chen. “*Teacher-Student Architecture for Knowledge Learning: A Survey*”. arXiv preprint arXiv:2210.17332.

Journal Articles

- J6 Shangyu Wu, Ying Xiong, Yufei Cui, **Haolun Wu**, Can Chen, Ye Yuan, Lianming Huang, Xue Liu, Tei-Wei Kuo, Nan Guan, Chun Jason Xue. “*Retrieval-Augmented Generation for Natural Language Processing: A Survey*”. Artificial Intelligence Review, 2026.
- J5 Ye Yuan, Youyuan Zhang, Can Chen, **Haolun Wu**, Zixuan Li, Jianmo Li, James J. Clark, and Xue Liu. “*Design Editing for Offline Model-based Optimization*”. Transactions on Machine Learning Research, 2025.
- J4 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, and Jiangchuan Liu. “*Large Language Model (LLM) for Telecommunications: A Comprehensive Survey on Principles, Key Techniques, and Opportunities*”. Communications Surveys and Tutorials, 2024.
- J3 **Haolun Wu***, Yansen Zhang*, Chen Ma, Fuyuan Lyu, Bhaskar Mitra, and Xue Liu. “*Result Diversification in Search and Recommendation: A Survey*”. IEEE Transactions on Knowledge and Data Engineering, 2024.
- J2 Chang Meng, Ziqi Zhao, Wei Guo, Yingxue Zhang, **Haolun Wu**, Chen Gao, Dong Li, Xiu Li, Ruiming Tang. “*Coarse-to-Fine Knowledge-Enhanced Multi-Interest Learning Framework for Multi-Behavior Recommendation*”. ACM Transactions on Information Systems, 2023.
- J1 **Haolun Wu**, Chen Ma, Bhaskar Mitra, Fernando Diaz, and Xue Liu. “*A Multi-objective Optimization Framework for Multi-stakeholder Fairness-aware Recommendation*”. ACM Transactions on Information Systems, 2022.

Conference Proceedings

- C31 Ye Yuan, Rui Song, Weien Li, Zeyu Li, Haochen Liu, Xiangyu Kong, Changjiang Han, Yonghan Yang, Zichen Zhao, Zixuan Dong, Fuyuan Lyu, Bowei He, **Haolun Wu**, Jikun Kang, Xue Liu. “*QUACK: Questioning, Understanding, and Auditing Communicated Knowledge in Multimodal Social Deduction Agents*”. In the 21st Conference on Empirical Methods in Natural Language Processing. Budapest, Hungary. Oct. 2026.

- C30 Yonghan Yang, Ye Yuan, Zipeng Sun, Linfeng Du, Bowei He, **Haolun Wu**, Can Chen, and Xue Liu. “*Support-ICML Proximity Augmented Diffusion Estimation for Offline Black-Box Optimization*”. In the 43rd International Conference on Machine Learning (ICML). Seoul, South Korea. Jul. 2026.
Poster, 26.6%
- C29 Zipeng Sun, Can Chen, Ye Yuan, **Haolun Wu**, Jiayao Gu, Christopher Pal, and Xue Liu. “*Training Diffusion-ICML Language Models for Black-Box Optimization*”. In the 43rd International Conference on Machine Learning (ICML). Seoul, South Korea. Jul. 2026.
Spotlight, 2.2%
- C28 Weixu Zhang, Ye Yuan, Changjiang Han, Yuxing Tian, Zipeng Sun, Linfeng Du, Jikun Kang, Hong Kang, ACL Xue Liu, **Haolun Wu***. “*Preference Heads in Large Language Models: A Mechanistic Framework for Main Interpretable Personalization*”. In the 64th Annual Meeting of the Association for Computational Linguistics. San Diego, USA. Jul. 2026.
- C27 Difan Jiao, Yilun Liu, Ye Yuan, Zhenwei Tang, Linfeng Du, **Haolun Wu**, Ashton Anderson. “*LLM Safety-ACL From Within: Detecting Harmful Content with Internal Representations*”. In the 64th Annual Meeting of the Main Association for Computational Linguistics. San Diego, USA. Jul. 2026.
- C26 Linfeng Du, Ye Yuan, Zichen Zhao, Fuyuan Lyu, Emiliano Penaloza, Xiuying Chen, Zipeng Sun, Jikun Kang, ACL Laurent Charlin, Xue Liu, **Haolun Wu***. “*Optimizing User Profiles via Contextual Bandits for Retrieval- Main Augmented LLM Personalization*”. In the 64th Annual Meeting of the Association for Computational Linguistics. San Diego, USA. Jul. 2026.
- C25 Weixu Zhang, Fanghua Ye, Qiang Gao, Jian Li, **Haolun Wu**, Yuxing Tian, Sijing Duan, Nan Du, Xiaolong ACL Li. “*Context-Fidelity Boosting: Enhancing Faithful Generation through Watermark-Inspired Decoding*”. In the 64th Annual Meeting of the Association for Computational Linguistics. San Diego, USA. Jul. 2026.
Findings
- C24 Zixiao Wang, Duzhen Zhang, Juntian Zhang, Yuhao Liu, Guoming Li, **Haolun Wu**, Le Song, Xiuying Chen. ACL “*GlossaGen: Making Academic Translation Smarter with Glossing*”. In the 64th Annual Meeting of the Findings Association for Computational Linguistics. San Diego, USA. Jul. 2026.
- C23 Lingyao Li, Xiaoshan Huang, Renkai Ma, Ben Zefeng Zhang, **Haolun Wu**, Fan Yang, and Chen Chen. WWW “*LLM Use for Mental Health: Crowdsourcing Users’ Sentiment-based Perspectives and Values from Social-20.1% Discussions*”. In the 35th Web Conference (WWW). Dubai, UAE. Apr. 2026.
- C22 Xiangwen Wang, Yibo Jacky Zhang, Zhoujie Ding, Katherine Tsai, **Haolun Wu**, Sanmi Koyejo. “*Aligning-NeurIPS Compound AI Systems via System-level DPO*”. In the 39th Annual Conference on Neural Information Processing Systems (NeurIPS). San Diego, USA. Dec. 2025.
- C21 Xiaoshan Huang, Tianlong Zhong, **Haolun Wu**, Yeyu Wang, Ethan Churchill, Xue Liu, David Williamson CSCW Shaffer. “*Linking Facial Recognition of Emotions and Socially Shared Regulation in Medical Simulation*”. In the 28th ACM SIGCHI Conference on Computer-Supported Cooperative Work & Social Computing (CSCW). Bergen, Norway. Oct. 2025.
- C20 Yansen Zhang, Bowei He, Xiaokun Zhang, **Haolun Wu**, Zexu Sun, Chen Ma. “*Counterfactual Multi-Player-ECML PKDD Bandits for Explainable Recommendation Diversification*”. In the European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases. Porto, Portugal. Sep. 2025.
- C19 Firat Öncel, Emiliano Penaloza, **Haolun Wu**, Shubham Gupta, Mirco Ravanelli, Laurent Charlin, Cem IEEE MLSP Subakan. “*Audio Prototypical Network for Controllable Music Recommendation*”. In the 35th IEEE International Workshop on Machine Learning for Signal Processing (MLSP). Istanbul, Turkey. Aug. 2025.
- C18 Fuyuan Lyu, Linfeng Du, Yunpeng Weng, Qiufang Ying, Zhiyan Xu, Wen Zou, **Haolun Wu**, Xiuqiang He, KDD Xing Tang. “*Timing is important: Risk-aware Fund Allocation based on Time-Series Forecasting*”. In the 31st ACM SIGKDD Conference on Knowledge Discovery and Data Mining. Toronto, Canada. Aug. 2025.
- C17 Gaurush Hiranandani*, **Haolun Wu***, Subhojyoti Mukherjee, and Sanmi Koyejo. “*Logits are All We Need-ICML to Adapt Closed Models*”. In the 42nd International Conference on Machine Learning (ICML). Vancouver, Canada. Jul. 2025.
Poster, 26.9%

- C16 Xiaoshan Huang, Jie Gao, **Haolun Wu**. “SSRLBot: Designing and Developing an LLM-based Agent using Socially Shared Regulated Learning”. In the 26th International Conference on Artificial Intelligence in Education (AIED). Palermo, Italy. Jul. 2025.
Oral, 24.2%
- C15 Xiaoshan Huang, **Haolun Wu**, Xue Liu, Susanne Lajoie. “What Makes Teamwork Work? A Multimodal Case Study on Emotions and Diagnostic Expertise in an Intelligent Tutoring System”. In the 26th International Conference on Artificial Intelligence in Education (AIED). Palermo, Italy. Jul. 2025.
Oral, 24.2%
- C14 Emiliano Penalosa, Olivier Gouvert, **Haolun Wu**, Laurent Charlin. “Text Representations for Scrutable Recommendations”. In the 34th Web Conference (WWW). Sydney, Australia. Apr. 2025.
The WebConf
Poster, 19.8%
- C13 **Haolun Wu**, Ofer Meshi, Masrour Zoghi, Fernando Diaz, Xue Liu, Craig Boutilier, and Maryam Karimzadehgan. “Density-based User Representation through Gaussian Process Regression for Multi-interest Personalized Retrieval”. In the 38th Annual Conference on Neural Information Processing Systems (NeurIPS). Vancouver, Canada. Dec. 2024.
NeurIPS
Poster, 25.8%
- C12 **Haolun Wu***, Ye Yuan*, Liana Mikaelyan, Alexander Meulemans, Xue Liu, James Hensman, and Bhaskar Mitra. “Learning to Extract Structured Entities Using Language Models”. In the 19th Conference on Empirical Methods in Natural Language Processing. Miami, Florida, USA. Nov. 2024.
EMNLP
Oral, top 3%
- C11 Ziqiang Cui, **Haolun Wu**, Bowei He, Ji Cheng, Kede Ma, Chen Ma. “Diffusion-based Contrastive Learning for Sequential Recommendation”. In the 33rd ACM International Conference on Information and Knowledge Management. Boise, Idaho, USA. Oct. 2024.
CIKM
Oral, 23.0%
- C10 **Haolun Wu**, Bhaskar Mitra, Nick Craswell. “Towards Group-aware Search Success”. In the 10th ACM SIGIR / the 14th International Conference on the Theory of Information Retrieval. Washington DC, USA. Jul. 2024.
SIGIR ICTIR
Oral, 41.0%
- C9 Xiaoshan Huang*, **Haolun Wu***, Xue Liu, and Susanne Lajoie. “Examining the Role of Peer Acknowledgements on Social Annotations: Unraveling the Psychological Underpinnings”. In the 41st ACM Conference on Human Factors in Computing Systems. Honolulu, USA. May. 2024.
CHI
Oral, 26.4%
- C8 **Haolun Wu***, Chengming Hu*, Xuan Li, Chen Ma, Xi Chen, Boyu Wang, Jun Yan, and Xue Liu. “Less or More from Teacher: Exploiting Trilateral Geometry for Knowledge Distillation”. In the 12th International Conference on Learning Representations. Vienna, Austria. May. 2024.
ICLR
Poster, 30.8%
- C7 **Haolun Wu**, Chen Ma, Yingxue Zhang, Xue Liu, and Mark Coates. “Self-supervised Contrastive Alignment for Tag-enhanced Recommendation”. For the 39th IEEE International Conference on Data Engineering. Anaheim, USA. Apr. 2023.
ICDE
Oral, 30.6%
- C6 **Haolun Wu**, Chen Ma, Yingxue Zhang, Xue Liu, and Mark Coates. “Adapting Triplet Importance of Implicit Feedback for Personalized Recommendation”. In the 31st ACM International Conference on Information and Knowledge Management. Atlanta, USA. Oct. 2022.
CIKM
Oral, 21.3%
- C5 **Haolun Wu***, Bhaskar Mitra*, Chen Ma, Fernando Diaz, and Xue Liu. “Joint Multisided Exposure Fairness for Recommendation”. In the 45th International ACM SIGIR Conference on Research and Development in Information Retrieval. Madrid, Spain. Jul. 2022.
SIGIR
Oral, 20.3%
- C4 Chen Ma, Liheng Ma, Yingxue Zhang, **Haolun Wu**, Xue Liu, and Mark Coates. “Knowledge-Enhanced Top-K Recommendation in Poincaré Ball”. In the 35th AAAI Conference on Artificial Intelligence. Virtual. Feb. 2021.
AAAI
Oral, 21.4%
- C3 **Haolun Wu**, Yunfei Feng. “A Pioneering Scalable Self-driving Car Simulation Platform”. In the IEEE 2nd International Conference on Automation, Electronics and Electrical Engineering. Shenyang, China. Nov. 2019.
AUTEEE
- C2 Chong Fu, Yufu Shan, Muiyang He, Ziyuan Yu, and **Haolun Wu**. “A New Medical Image Encryption Algorithm Using Multiple 1-D Chaotic Maps”. In the IEEE International Conference on Systems, Man, and Cybernetics. Miyazaki, Japan. Oct. 2018.
SMC

- C1 Yufu Shan, Muiyang He, Ziyuan Yu, and **Haolun Wu**. “Pixel Level Image Encryption Based on Semantic Segmentation”. In the IEEE International Conference on Control, Artificial Intelligence, Robotics and Optimization. Prague, Czech Republic. May. 2018.

Workshops

- W3 Ye Yuan, Youyuan Zhang, Can Chen, **Haolun Wu**, Melody Zixuan Li, Jianmo Li, James J. Clark, Xue Liu.
ICLR workshop “Design Editing for Offline Model-based Optimization”. In ICLR 2025 Workshop on Deep Generative Model in Machine Learning: Theory, Principle and Efficacy.
- W2 Fuyuan Lyu, Linfeng Du, Yunpeng Weng, Qiufang Ying, Zhiyan Xu, wenzou, **Haolun Wu**, Xiuqiang He,
ICLR workshop Xing Tang. “Timing is important: Risk-aware Fund Allocation based on Time-Series Forecasting”. In ICLR 2025 Workshop on Financial AI.
- W1 **Haolun Wu**, Ofer Meshi, Masrour Zoghi, Fernando Diaz, Xue Liu, Craig Boutilier, and Maryam Karimzade-
WSDM hgan. “Density-based User Representation through Gaussian Process Regression for Multi-interest Personal-
workshop ized Retrieval”. In WSDM 2024 Workshop on Learning Representation and Clustering.

Patents

- P3 **Haolun Wu**, Chen Ma, Yingxue Zhang, and Mark Coates. “Recommendation System With Adaptive Thresholds for Neighborhood Selection”. US 2022/0253722 A1. Patent US 2022.
- P2 **Haolun Wu**, Chen Ma, and Yingxue Zhang. “Recommendation System with Adaptive Weighted Bayesian Personalized Ranking Loss”. US 2022/0253688 A1. Patent US 2022.
- P1 **Haolun Wu**, Haodong Lai, Yutao Liu, Wentao Zhou, and Qinlai Li. “A Method For Real-Time Image Style Transfer Based On Conditional Generative Adversarial Networks ”. AU2017101166A4. IP Australia 2017.

Grants & Fellowships

Total funding: \$484,334. Total share of funding: \$269,667.

I led or made core contribution to the writing of the following grant and fellowship proposals during Ph.D.

- 2024 [Fonds de recherche du Québec](#) FRQNT Ph.D. Fellowship (\$58,334, share 100%) (rank 1st-place)
- 2024 [Samsung Global Research Outreach Program](#) (\$205,000, share 30%) (1 of the 11 recipients worldwide)
- 2023 [Borealis AI Fellowship](#) (\$10,000, share 100%) (1 of the 10 recipients in Canada)
- 2022 MSR-Mila Collaboration Grant (\$66,000, share 50%)
Lead PI: Xue (Steve) Liu; Industrial Mentor: Bhaskar Mitra, John Winn
Title: SNAKE: Structured Neural Attention over Knowledge base Entities
Funding Body: Microsoft Research
- 2022 MITACS Accelerate Fellowship (\$90,000, share 50%)
Lead PI: Xue (Steve) Liu; Industrial Mentor: Shuo Wang
Title: Fairness-aware Matching for Dynamic Predictive Behavior Matching
Funding Body: Bell Canada, MITACS
- 2020 MSR-Mila Collaboration Grant (\$55,000, share 100%)
Lead PI: Xue (Steve) Liu; Industrial Mentor: Fernando Diaz
Title: [Two-sided Recommendation with Fairness](#)
Funding Body: Microsoft Research

Honors & Awards

Total award amount: \$58,800.

- 2026 [Governor General's Gold Medal](#), *McGill University & Governor General of Canada*
- 2026 D. W. Ambridge Prize (\$1,500), *McGill University*
- 2026 Tinker Research Grant (\$5,000), *Thinking Machine Lab*
- 2026 Tinker Research Grant (\$5,000), *Thinking Machine Lab*
- 2026 Cohere Labs Catalyst Grant (\$1,000), *Cohere Labs*
- 2024 Student Travel Grant (\$3,000), *Neural Information Processing Systems (NeurIPS)*
- 2024 Grad Excellence Award (\$6,000), *McGill University*
- 2024 Student Travel Award (\$1,000), *International Conference on Learning Representations (ICLR)*
- 2023 Grad Excellence Award (\$6,000), *McGill University*
- 2023 Graduate Stimulus and Financial Supplement (\$1,750+\$1,500+\$850), *McGill University*
- 2022 Grad Excellence Award (\$7,000+\$1,000), *McGill University*
- 2022 SIGIR Student Travel Grant. (\$1,350+\$850)
- 2021 Apple Scholars PhD Fellowship Nominee (1 out of 3 students at *McGill University*)
- 2021 Grad Excellence Award (\$5,000), *McGill University*
- 2020 Grad Excellence Award (\$12,600), *McGill University*
- 2019 Excellent Graduation Thesis, *Northeastern University*
- 2016-2019 Outstanding Student (\$2,400), *Northeastern University*
- 2017 First Prize, *CUMCM (Contemporary Undergraduate Mathematical Contest in Modelling)*
- 2017 MCM/ICM Meritorious Winner, *COMAP (Consortium for Mathematics and Its Application)*
- 2016 Yao Tianshun Scholarship (\$1,000, 2 out of 250 students), *Northeastern University*
- 2014 First Prize (rank 18th in [Sichuan Province](#)), *CMS (Chinese Mathematical Association)*

Media

- 2023 [McGill PhD Candidates Earn Borealis AI Fellowships](#). McGill Reporter.

Community & Leadership

o Mila Lab Representatives

- I am honored to have been elected and to serve as one of the [lab representatives \(11 out of 1000+ students\)](#) for 2023-24 at *Mila - Quebec AI Institute*, where I proudly represent the McGill PhD cohort. The Mila LabReps act as a bridge between students and the various members of Mila's ecosystem. I am actively involved in organizing student assemblies, writing newsletters, managing finances, and facilitating communication among students, professors, and staff.

o Founder, Organizer, Core Contributor

- I am a **co-founder and organizer** of the [DEFIRST](#) reading group at *Mila* and the *Vector Institute* to build an interdisciplinary forum of researchers across both industry and academia, who work at the intersection of Information Retrieval, Fairness, Generative Model, and Trustworthy AI. Past talks are on our YouTube: <https://www.youtube.com/@defirstreadinggroup-milaxv8747/videos>.
- I am also a **core steering committee member** for **NICE (NLP Academic Exchange Platform)** in both [Chinese version](#) and [English version](#). NICE is a large and fast-growing community dedicated to promoting open exchange, collaboration, and frontier research discussion in NLP. NICE now reaches thousands of followers, shares weekly NLP research highlights, and brings together researchers from academia and industry. To date, it has now organized 150+ academic events, hosted 300+ invited speakers from around the world, accumulated 150K+ followers across platforms, and garnered 1M+ total video views. I help curate research content, organize community events, and facilitate cross-institutional discussions, contributing to a platform that meaningfully improves visibility and engagement across the NLP research community.
- I also serve as a **core steering Committee member (Academic Track)** for the vLLM Semantic Router (vSR), an

open-source project in the vLLM ecosystem building a programmable Mixture-of-Models router for heterogeneous LLM inference. Since its launch in September 2025, vSR has achieved substantial real-world adoption, surpassing one million downloads collectively across Hugging Face (350K+), PyPI (450K+), and GitHub (200K+), drawing 50+ contributing researchers and engineers worldwide, and being adopted by companies including Microsoft, AMD, Red Hat, UBS, and Bayer, with public recognition from industry leaders including Microsoft CEO Satya Nadella and Hugging Face CEO Clément Delangue. On the academic track, I help shape the project's research agenda across semantic routing, reasoning-effort allocation, and safety-aware model selection, coordinate benchmarking and publication efforts. This role lets me translate frontier research into infrastructure deployed at scale, contributing to a project that meaningfully improves the efficiency and safety of large-scale LLM inference.

- I am one of the **organizers** of the [OracleLLM: Empowering LLM with Self-Feedback](#) community. OracleLLM refers to replacing human-generated golden references with the output and feedback produced by LLMs, treating them as authoritative benchmarks and learning signals for various tasks and applications.
- I am one of the **significant contributors** to the [Test-Time Scaling \(TTS\)](#) survey and community, a collaborative effort to understand and push the boundaries of how LLMs can adapt and improve at inference time.

○ Participant

- [Summer School in Responsible Artificial Intelligence \(AI\) and Human Rights](#) hosted by Mila and University of Montreal (selected as 1 of the 40 global participants).

○ Session Chair

- SIGIR 2022 (Session: Collaborative Filtering)

○ PC member / Reviewer

- **2026:** NeurIPS, ICLR, ICML, KDD, AAAI, TKDD, TKDE, SIGIR, WWW, ARR Rolling Review
- **2025:** NeurIPS, ICLR, ICML, KDD, AAAI, TKDD, TKDE, SIGIR, WWW, ARR Rolling Review
- **2024:** NeurIPS, ICLR, ICML, KDD, AAAI, TKDD, TKDE, SIGIR, WWW, ARR Rolling Review
- **2023:** AAAI, TKDD, TKDE, SIGIR
- **2022:** AAAI, TKDD, SIGIR

————— Talk, Lecture & Mentorship

○ Talks and Lectures

- **Talk at MBZUAI:** *Two Pathways to Trustworthy AI: Transparent Representations and Robust Adaptation*. [Rising-Star Talk, International Symposium on Trustworthy Foundation Models 2025](#). May 2025.
- **Talk at Google:** *Two Pathways to Trustworthy AI: Transparent Representations and Robust Adaptation*. May 2025.
- **Talk at Stanford University:** *Human-Centric Alignment through Multilevel Feedback Integration*. Mar. 2025.
- **Lecture at McGill University:** *Applications of Machine Learning in Real-World Systems (COMP 597/598)*. Led weekly discussion sections and instructional activities. Winter 2021 & 2022.
- **Talk at Microsoft Research:** *Multi-FR: A Multi-Objective Optimization Method for Achieving Two-sided Fairness in E-commerce Recommendation*. Oct. 2021.
- **Talk at Mila:** *Joint Multisided Exposure Fairness for Search and Recommendation*. Sep. 2022. [Selected as 1 of 6 Contributed Talks, Montreal AI Symposium 2022](#).
- **Talk at Bell Canada:** *Joint Multisided Exposure Fairness for Search and Recommendation*. Aug. 2022.
- **Talk at Microsoft Research:** *Joint Multisided Exposure Fairness for Search and Recommendation*. Apr. 2022.

○ Mentored eight undergraduate and graduate students. (2022-)

[The following mentorship leads to 6 papers accepted to ACL, EMNLP, WWW, ICML \(including one spotlight\), and helps two students get their graduate offers from the University of Toronto and the University of Waterloo.](#)

- | | | |
|-------------------------------------|----------------------|-----------------------------|
| - Hillary Tao (McGill) | - Ye Yuan (Mila) | - Harry Yang (MBZUAI) |
| - Yansen Zhang (CityU of Hong Kong) | - Weixu Zhang (Mila) | - Difan Jiao (U of Toronto) |
| - Rui Song (McGill) | - Linfeng Du (Mila) | - Changjiang Han (MBZUAI) |
| - Evelyn Cao (U of Toronto) | - Zipeng Sun (Mila) | |

Coding Skills

Python, Pytorch, Tensorflow, Jax.

Extracurricular Activities

- I am a certificated **clarinet player**. I was a member of Musicians Association of Sichuan, China. My advisor was Yi Zhu, who is now a Deputy Director of the Music Department of Sichuan University.
- I am also a master in **Chinese Calligraphy**. My advisor was Jialin Zheng, who is now a Vice Chairman of Sichuan Calligraphers Association and Vice Chairman of Chengdu Calligraphers Association.

Referees

- **Steve Liu**

Fellow of the Canadian Academy of Engineering (FCAE); Fellow of the IEEE (FIEEE)
Professor of Computer Science & Machine Learning, ex-Associate VP of Research, *MBZUAI, Abu Dhabi, UAE*
Professor & William Dawson Scholar, *School of Computer Science, McGill University, Montreal, Canada*
Associate Member, *Mila - Quebec AI Institute*
xue.liu@mcgill.ca

- **Craig Boutilier**

Principal Scientist, *Google Research, Mountain View, USA*
ex-Professor, *Department of Computer Science, University of Toronto, Toronto, Canada*
cboutilier@google.com

- **Fernando Diaz**

Associate Professor, *Language Technologies Institute, Carnegie Mellon University, Pittsburgh, USA*
Research Scientist, *Google Research, Montreal, Canada*
diazf@acm.org

- **Sanmi Koyejo**

Assistant Professor, *Computer Science, Stanford University, Palo Alto, USA*
Stanford Trustworthy AI Research (STAIR), Stanford AI Laboratory (SAIL), Stanford Human-centered AI (HAI)
sanmi@cs.stanford.edu