MINSU KIM

Postdoctoral Fellow
Mila - Quebec AI Institute

✉ minsu.kim@mila.quebec
G google scholar
🌐 homepage

HIGHLIGHTS

- 40+ peer-reviewed papers, incl. 23 NeurIPS/ICML/ICLR (10 lead); 14 lead overall

- Joint projects with LLNL, Intel, Samsung, LG, KT, and LawZero, bridging research and industry applications

- Research areas:

  - **ML fundamentals**: exploration, credit assignment, probabilistic inference, generative modeling, distributed RL
  - **RL for post-training**: aligning LLMs, diffusion, GANs, VAEs, protein folding
  - **Trustworthy reasoning of LLMs**: red-teaming, uncertainty quantification, self-verification
  - **AI4Opt**: neural solvers for NP-hard problems, RL for multi-robot distributed scheduling
  - **AI4Science**: generative modeling and RL for molecules, 3D conformations, biological sequences
  - **AI4Sys**: ML-driven optimization for physical systems, including HBM, interconnects, and power distribution networks

EMPLOYMENT AND COLLABORATIONS

6/2025 – Present   **Postdoctoral Fellow (Host: Prof. Yoshua Bengio)**   **Mila**
- Trustworthy reasoning in LLM
- **Joint affiliation with KAIST via the KAIST–Mila Prefrontal Research Collaboration (Profs. Sungjin Ahn and Sungsoo Ahn)*

6/2025 – Present   **Academic Collaborator**   **LawZero**
- Led collaboration on trustworthy LLMs; focusing on red-teaming and safe reasoning
- Contributed to the uncertainty measurement of LLMs project with joint experiments and publications

1/2025 – 6/2025
12/2023 – 5/2024   **Research Intern (Supervisor: Prof. Yoshua Bengio)**   **Mila**
- GFlowNets for LLM/LMM fine-tuning
- GFlowNets for biological and chemical discovery

EDUCATION

3/2022 – 2/2025   **Ph.D. in Industrial and Systems Engineering (Advisor: Prof. Jinkyoo Park)**   **KAIST**
Thesis: Off-policy training methods for probabilistic agents in combinatorial space
<u>**Presidential Best Ph.D. Thesis Award**</u>

3/2020 – 2/2022   **M.S. in Electrical Engineering (Advisor: Prof. Joungho Kim)**   **KAIST**
Research: Applied reinforcement learning to routing problems in high-speed interconnection design

3/2015 – 2/2020   **B.S. in Mathematics and Computer Science (Dual Major)**   **KAIST**

INDUSTRY-ACADEMIA COLLABORATION PROJECTS

2025 – Present   **Scalable Multi-agent Off-policy Distributed Training of GFlowNets**   **Intel–Mila**
- Research on distributed training pipelines and analyzing scaling laws on multi-GPU clusters

2024 – 2025   **Distributed Off-Policy RL for Efficient LLM Post-Training**   **Lawrence Livermore National Lab (LLNL) - Mila**
- Co-authored work (NeurIPS 2025) on asynchronous off-policy RL to improve LLM fine-tuning efficiency
- Achieved 4× faster wall-clock time in LLM post-training via asynchronous off-policy RL, enabling scalable deployment

2022 – 2023   **Multi-Robot Distributed Scheduling with Graph RL**   **KT - KAIST**
- Cluster-robot scheduling using reinforcement learning on graphs
- Led research publication of the project (AAAI 2024)

2020 – 2021   **AI for Semiconductor Design Automation considering Signal & Power Integrity**   **Samsung-KAIST**
- Applied ML in high-speed interconnect design
- Led research publication of the project (EPEPS 2022)

TOP 10 SELECTED PUBLICATIONS
(see THIS LINK for full publication)

Preprint, 2025	**Active Attacks: Redteaming LLMs via Adaptive Environments** Taeyoung Yun, Pierre-Luc St-Charles, Jinkyoo Park, Yoshua Bengio, and Minsu Kim
Preprint, 2025	**Latent Veracity Inference for Identifying Errors in Stepwise Reasoning** {Minsu Kim*, Jean-Pierre Falet*}, Oliver E. Richardson, Xiaoyin Chen, Moksh Jain, Sungjin Ahn, Sungsoo Ahn, and Yoshua Bengio
NeurIPS, 2025	**On Scalable and Efficient Training of Diffusion Samplers** Minkyu Kim, Kiyoung Seong, Dongyeop Woo, Sungsoo Ahn, and Minsu Kim
ICLR, 2025	**Adaptive Teachers for Amortized Samplers** {Minsu Kim*, Sanghyeok Choi*}, Taeyoung Yun, Emmanuel Bengio, Leo Feng, Jarrid Rector-Brooks, Sungsoo Ahn, Jinkyoo Park, Nikolay Malkin, and Yoshua Bengio
NeurIPS, 2024	**Amortizing Intractable Inference in Diffusion Models for Vision, Language, and Control** {Siddarth Venkatraman*, Moksh Jain*, Luca Scimeca*, Minsu Kim*, Marcin Sendera*} et al.
ICML 2024	**Learning to Scale Logits for Temperature-Conditional GFlowNets** {Minsu Kim*, Joohwan Ko*, Taeyoung Yun*}, Dinghuai Zhang, Ling Pan, Taeyoung Yun, Woochang Kim, Jinkyoo Park, Emmanuel Bengio, and Yoshua Bengio
ICLR, 2024	**Local Search GFlowNets** Minsu Kim, Taeyoung Yun, Emmanuel Bengio, Dinghuai Zhang, Yoshua Bengio, Sungsoo Ahn, and Jinkyoo Park
NeurIPS, 2023	**Bootstrapped Training of Score-Conditioned Generator for Offline Design of Biological Sequences** Minsu Kim, Federico Berto, Sungsoo Ahn, and Jinkyoo Park
NeurIPS, 2022	**Sym-nco: Leveraging Symmetricity for Neural Combinatorial Optimization** Minsu Kim, Junyoung Park, and Jinkyoo Park
NeurIPS, 2021	**Learning Collaborative Policies to Solve NP-hard Routing Problems** Minsu Kim, Jinkyoo Park, and Joungho Kim

SELECTED HONORS AND AWARDS (LINK: FULL RECORDS)

6/2025	**CIFAR AI Safety Postdoctoral Fellowship (via Yoshua Bengio)** Awarded by the Canadian AI Safety Institute at CIFAR, under the mentorship of Prof. Yoshua Bengio	**CIFAR / Mila**
4/2025	**Jang Yeong Sil Fellowship Award** A prestigious awardrecognizing early-career postdoctoral researchers with outstanding potential	**KAIST**
12/2023	**Qualcomm Innovative Fellowship Award** Paper: Sym-NCO: Leveraging Symmetricity for Neural Combinatorial Optimization	**Qualcomm Korea**
2021	**Best Paper Award (First Author)** Paper: Neural Language Model Enables Extremely Fast and Robust Routing on Interposer	**DesignCon 2021**

INVITED TALKS AND INDUSTRY PRESENTATIONS

2025	**Rising Star Talk: Trustworthy Foundation Model**	**MBZUAI (Abu Dhabi)**
2024	**Invited Research Talk**	**Valence Labs (Virtual)**
2025	**Research Spotlight, LG Tech Day**	**LG AI Research (Seoul)**
2023	**Research Spotlight, Samsung AI Forum**	**Samsung Advanced Institute of Technology (Suwon)**