Bo Liu

+1 (347) 328-3747 | benjaminliu.eecs@gmail.com | Homepage | Google Scholar | GitHub | LinkedIn

Education

Ph.D. in Computer Science — Aug. 2023 – Present
School of Computing (SOC), National University of Singapore — *Singapore, Singapore*
Advisor: Prof. Wee Sun Lee and Prof. David Hsu

B.S. in Intelligence Science and Technology — Sep. 2016 – Jul. 2020
School of Electronics Engineering and Computer Science (EECS), Peking University — *Beijing, China*

B.Econ. in Economics — Sep. 2017 – Jul. 2020
National School of Development, Peking University — *Beijing, China*

Publications

* indicates equal contribution

Agent Learning via Early Experience
Kai Zhang, Xiangchao Chen*, **Bo Liu***, Tianci Xue*, Zeyi Liao*, Zhihan Liu*, Xiyao Wang, Yuting Ning, Zhaorun Chen, Xiaohan Fu, Jian Xie, Yuxuan Sun, Boyu Gou, Qi Qi, Zihang Meng, Jianwei Yang, Ning Zhang, Xian Li, Ashish Shah, Dat Huynh, Hengduo Li, Zi Yang, Sara Cao, Lawrence Jang, Shuyan Zhou, Jiacheng Zhu, Huan Sun, Jason Weston, Yu Su, Yifan Wu
International Conference on Machine Learning (**ICML**), 2026

SPIRAL: Self-Play on Zero-Sum Games Incentivizes Reasoning via Multi-Agent Multi-Turn Reinforcement Learning
Bo Liu, Leon Guertler, Simon Yu, Zichen Liu, Penghui Qi, Daniel Balcells, Mickel Liu, Cheston Tan, Weiyan Shi, Min Lin, Wee Sun Lee, Natasha Jaques
International Conference on Learning Representations (**ICLR**), 2026

Vision-Zero: Scalable VLM Self-Improvement via Strategic Gamified Self-Play
Qinsi Wang, **Bo Liu**, Tianyi Zhou, Jing Shi, Yueqian Lin, Yiran Chen, Hai Helen Li, Kun Wan, Wentian Zhao
International Conference on Learning Representations (**ICLR**), 2026

GEM: A Gym for Agentic LLMs
Zichen Liu, Anya Sims, Keyu Duan, Changyu Chen, Simon Yu, Xiangxin Zhou, Haotian Xu, Shaopan Xiong, **Bo Liu**, Chenmien Tan, Chuen Yang Beh, Weixun Wang, Hao Zhu, Weiyan Shi, Diyi Yang, Michael Shieh, Yee Whye Teh, Wee Sun Lee, Min Lin
International Conference on Learning Representations (**ICLR**), 2026

Scaling Agent Learning via Experience Synthesis
Zhaorun Chen, Zhuokai Zhao, Kai Zhang, **Bo Liu**, Qi Qi, Yifan Wu, Tarun Kalluri, Sara Cao, Yuanhao Xiong, Haibo Tong, Huaxiu Yao, Hengduo Li, Jiacheng Zhu, Xian Li, Dawn Song, Bo Li, Jason Weston, Dat Huynh
International Conference on Learning Representations (**ICLR**), 2026

Natural Language Reinforcement Learning
Xidong Feng*, **Bo Liu***, Ziyu Wan, Haotian Fu, Girish A. Koushik, Zhiyuan Hu, Mengyue Yang, Ying Wen, Jun Wang
International Conference on Learning Representations SSI-FM Workshop (**ICLR Workshop SSI-FM**), 2025

DeepSeek-Prover-V1.5: Harnessing Proof Assistant Feedback for Reinforcement Learning and Monte-Carlo Tree Search
Huajian Xin, Z. Z. Ren, Junxiao Song, Zhihong Shao, Wanjia Zhao, Haocheng Wang, **Bo Liu**, Liyue Zhang, Xuan Lu, Qiushi Du, Wenjun Gao, Qihao Zhu, Dejian Yang, Zhibin Gou, Z. F. Wu, Fuli Luo, Chong Ruan

International Conference on Learning Representations (**ICLR**), 2025

Differentiable Information Enhanced Model-Based Reinforcement Learning
Xiaoyuan Zhang, Xinyan Cai, **Bo Liu**, Weidong Huang, Song-Chun Zhu, Siyuan Qi, Yaodong Yang
AAAI Conference on Artificial Intelligence (**AAAI**), 2025 **Oral**

Grasp Multiple Objects with One Hand
Yuyang Li, **Bo Liu**, Yiran Geng, Puhao Li, Yaodong Yang, Yixin Zhu, Tengyu Liu, Siyuan Huang
IEEE Robotics and Automation Letters (**RA-L**), 2024
IEEE/RSJ International Conference on Intelligent Robots and Systems (**IROS**), 2024 **Oral**

TorchOpt: An Efficient Library for Differentiable Optimization
Jie Ren*, Xidong Feng*, **Bo Liu***, Xuehai Pan*, Yao Fu, Luo Mai, Yaodong Yang
Journal of Machine Learning Research (**JMLR**), 2023

A Theoretical Understanding of Gradient Bias in Meta-Reinforcement Learning
Bo Liu*, Xidong Feng*, Jie Ren, Luo Mai, Rui Zhu, Haifeng Zhang, Jun Wang, Yaodong Yang
Conference on Neural Information Processing Systems (**NeurIPS**), 2022

EnvPool: A Highly Parallel Reinforcement Learning Environment Execution Engine
Jiayi Weng, Min Lin, Shengyi Huang, **Bo Liu**, Denys Makoviichuk, Viktor Makoviychuk, Zichen Liu, Yufan Song, Ting Luo, Yukun Jiang, Zhongwen Xu, Shuicheng Yan
Conference on Neural Information Processing Systems Datasets and Benchmarks (**NeurIPS D&B**), 2022

Neural Auto-Curricula in Two-Player Zero-Sum Games
Xidong Feng, Oliver Slumbers, Ziyu Wan, **Bo Liu**, Stephen McAleer, Ying Wen, Jun Wang, Yaodong Yang
Conference on Neural Information Processing Systems (**NeurIPS**), 2021

Learning Correlated Communication Topology in Multi-Agent Reinforcement Learning
Yali Du, **Bo Liu**, Vincent Moens, Ziqi Liu, Zhicheng Ren, Jun Wang, Xu Chen, Haifeng Zhang
International Conference on Autonomous Agents and Multiagent Systems (**AAMAS**), 2021 **Oral**

Preprints

* indicates equal contribution

Agents' Last Exam
Yiyou Sun, Xinyang Han, Weichen Zhang, Yuanbo Pang, Tianyu Wang, Yuhan Cao, Yixiao Huang, Chris Duroiu, Haoyun Zhang, Jeffrey Lin, Weishu Zhang, Tyler Zeng, Ying Yan, **Bo Liu**, Hanson Wen, Mingyang Xu, Xiaoyuan Liu, Zimeng Chen, Weiyan Shi, Amanda Dsouza, Vincent Sunn Chen, Dawn Song, and others
Preprint on **arXiv**, 2026

SPICE: Self-Play In Corpus Environments Improves Reasoning
Bo Liu, Chuanyang Jin, Seungone Kim, Weizhe Yuan, Wenting Zhao, Ilia Kulikov, Xian Li, Sainbayar Sukhbaatar, Jack Lanchantin, Jason Weston
Preprint on **arXiv**, 2025

BigCodeArena: Unveiling More Reliable Human Preferences in Code Generation via Execution
Terry Yue Zhuo, Xiaolong Jin, Hange Liu, Juyong Jiang, Tianyang Liu, Chen Gong, Bhupesh Bishnoi, Vaisakhi Mishra, Marek Suppa, Noah Ziems, Saiteja Utpala, Ming Xu, Guangyu Song, Kaixin Li, Yuhan Cao, **Bo Liu**, Zheng Liu, Sabina Abdurakhmanova, Wenhao Yu, Mengzhao Jia, Jihan Yao, Kenneth Hamilton, Kumar Shridhar, Minh Chien Vu, Dingmin Wang, Jiawei Liu, Zijian Wang, Qian Liu, Binyuan Hui, Meg Risdal, Ahsen Khaliq, Atin Sood, Zhenchang Xing, Wasi Uddin Ahmad, John Grundy, David Lo, Banghua Zhu, Xiaoning Du, Torsten Scholak, Leandro von Werra
Preprint on **arXiv**, 2025

The Era of Real-World Human Interaction: RL from User Conversations
Chuanyang Jin, Jing Xu*, **Bo Liu***, Leitian Tao, Olga Golovneva, Tianmin Shu,
Wenting Zhao, Xian Li, Jason Weston
Preprint on **arXiv**, 2025

LLaVA-Critic-R1: Your Critic Model is Secretly a Strong Policy Model
Xiyao Wang, Chunyuan Li, Jianwei Yang, Kai Zhang, **Bo Liu**, Tianyi Xiong, Furong Huang
Preprint on **arXiv**, 2025

TextArena
Leon Guertler, Bobby Cheng, Simon Yu, **Bo Liu**, Leshem Choshen, Cheston Tan
Preprint on **arXiv**, 2025

DeepSeek-Prover: Advancing Theorem Proving in LLMs through Large-Scale Synthetic Data
Huajian Xin, Daya Guo, Zhihong Shao, Zhizhou Ren, Qihao Zhu, **Bo Liu**,
Chong Ruan, Wenda Li, Xiaodan Liang
Preprint on **arXiv**, 2024

DeepSeek-V2: A Strong, Economical, and Efficient Mixture-of-Experts Language Model
DeepSeek-AI: **Bo Liu**, and others (alphabetic order)
Preprint on **arXiv**, 2024

DeepSeek-VL: Towards Real-World Vision-Language Understanding
Haoyu Lu, Wen Liu, Bo Zhang, Bingxuan Wang, Kai Dong, **Bo Liu**, Jingxiang Sun, Tongzheng Ren,
Zhuoshu Li, Hao Yang, Yaofeng Sun, Chengqi Deng, Hanwei Xu, Zhenda Xie, Chong Ruan
Preprint on **arXiv**, 2024

DeepSeek-LLM: Scaling Open-Source Language Models with Longtermism
DeepSeek-AI: **Bo Liu**, and others (alphabetic order)
Preprint on **arXiv**, 2024

Selected Research Projects

- **University of Washington** Dec. 2025 – Present
  Visiting Researcher, Advisor: Natasha Jaques *Seattle, WA*

  SPIRAL: Self-Play on Zero-Sum Games Incentivizes Reasoning via Multi-Agent Multi-Turn Reinforcement Learning
  - Developed SPIRAL, a self-play RL framework where LLMs learn reasoning by playing multi-turn zero-sum games against continuously improving versions of themselves, removing reliance on human-curated problem-answer pairs and domain-specific reward engineering
  - Designed Role-conditioned Advantage Estimation (RAE) to stabilize multi-agent policy gradients and prevent chain-of-thought reasoning-trace collapse unique to LLM self-play training
  - Implemented a fully online, multi-turn, multi-agent RL system for LLMs with REINFORCE-based policy optimization across three zero-sum games (TicTacToe, Kuhn Poker, Simple Negotiation) selected for cognitive diversity
  - Demonstrated up to 10% improvement across 8 reasoning benchmarks on 4 models spanning the Qwen and Llama families, outperforming SFT on 25,000 expert game trajectories, with RLVR-trained models (e.g., DeepSeek-R1-Distill-Qwen-7B) further benefiting from SPIRAL as a complementary booster
  - Completed a co-first-authored paper published at **ICLR 2026**

- **Meta FAIR** May 2025 – Nov. 2025
  Research Scientist Intern, Advisor: Jason Weston *New York, NY*

  SPICE: Self-Play In Corpus Environments Improves Reasoning
  - Developed SPICE, a novel self-play RL framework where a single model acts as both Challenger (generating corpus-grounded tasks) and Reasoner (solving tasks), achieving consistent gains on various reasoning benchmarks

- Designed variance-based curriculum reward system that creates adaptive challenge at the frontier of model capability, enabling co-evolution of question generation and solving abilities
- Implemented distributed actor-learner architecture using DrGRPO with role-specific advantages for joint optimization of adversarial dynamics between Challenger and Reasoner
- Demonstrated that corpus grounding prevents hallucination amplification and information symmetry issues plaguing pure self-play methods, achieving consistent gains across mathematical (+8.9%) and general reasoning (+9.8%) tasks
- Completed a first-authored paper published at **arXiv 2025**

- **DeepSeek** Sep. 2023 – Sep. 2024
Student Researcher *Beijing, China*

Foundation Model Development and Post-Training

- Built RLHF infrastructure (HAI-Chat v1.0) for DeepSeek-LLM/V2, transitioning from DeepSpeed to Megatron with 3D parallelism for PPO/DPO at 236B MoE scale
- Proposed removing critic network from PPO, significantly improving training stability and efficiency, contributing to the development of GRPO now used in DeepSeek-R1
- Designed and curated RLHF training data pipeline used in DeepSeek-Coder, ensuring high-quality preference datasets for alignment of large-scale models
- Led DeepSeek-VL post-training via self-questioning: model generates questions from images and chain-of-thought answers for self-improvement vision-language learning
- Architected event-based Lean server for DeepSeek-Prover/V1.5 processing thousands of outputs simultaneously, achieving 63.5% on miniF2F with MCTS and iterative DPO
- Core contributor to 4 models: **DeepSeek-LLM**, **DeepSeek-V2**, **DeepSeek-VL**, and **DeepSeek-Prover-V1.5**

- **Microsoft Research Asia** Jun. 2023 – Sep. 2023
Research Intern, Advisor: Yaobo Liang *Beijing, China*

AutoLearner: Autonomously Learning Explicit Knowledge from Environments

- Proposed the initial version of the AutoLearner framework, which contains two stages: the agent autonomously generates goals and collects data during the completion of these goals; in the execution stage, the agent utilizes the goal skill learning knowledge to accomplish new goals effectively
- Developed the initial version of the AutoLearner framework utilizing a LLM world model for enhanced planning and decision-making during goal achievement
- Demonstrated the efficacy of our method across various OS tasks, showing that the acquired knowledge substantially enhances performance and can be generalized to new environments, and transferred across different LLMs and agents
- Completed a co-authored paper submitted to **NeurIPS 2024**

- **Peking University** Mar. 2023 – Jun. 2023
Research Assistant, Advisor: Yaodong Yang *Beijing, China*

Grasp Multiple Objects with One Hand

- Formulated the first Goal-Conditioned Reinforcement Learning (GCRL) policy dedicated to the simultaneous grasping and lifting of multiple objects; implemented RL training infrastructure for training and reward shaping
- Adapted the IBS-Grasping method for multi-object scenarios to provide a comparative baseline
- Completed a second-authored paper published at **RA-L 2024** and at **IROS 2024 Oral**

- **University College London** Jul. 2020 – Mar. 2023
Research Assistant, Advisor: Jun Wang *Beijing, China*

A Theoretical Understanding of Gradient Bias in Meta-Reinforcement Learning

- Proposed a unified framework for variations of gradient-based meta-reinforcement learning (GMRL) algorithms
- Derived upper bound for two newly identified biases in GMRL: compositional bias and multi-step hessian bias
- Conducted ablation studies qualitatively and quantitatively to verify how the bias terms affect estimation quality
- Designed two plug-and-play methods for bias mitigation: off-policy learning correction and LVC Hessian estimator
- Completed a co-first-authored paper published at **NeurIPS 2022**

Learning Correlated Communication Topology in Multi-Agent Reinforcement Learning

- Proposed FlowComm to evaluate the correlation between agent communication interactions and learned message-augmented decentralized policies & graph reasoning policies through multi-agent reinforcement learning
- Generalized coupling flow to model the interaction graph of MARL conditioning on the global states of all agents
- Demonstrated through visualizations on Particle World that FlowComm learned meaningful communications
- Completed a second-authored paper published at **AAMAS 2021**

- **Carnegie Mellon University** Oct. 2019 – Jan. 2020
  Undergraduate Research Assistant, Advisor: Changliu Liu *Pittsburgh, PA*

  Multi-UAV Collaborative Transportation
  - Proposed a hierarchical training approach to explore cooperative strategy for multi-UAV collaborative transportation
  - Established RL environment that combined OpenAI gym with real world UAV models within ROS Gazebo simulator
  - Developed a modified MADDPG algorithm based on Bidirectional LSTM as centralized value function and then combined it with low-level control to identify the optimal strategy for UAVs
  - Designed a novel RL learning agent to assign appropriate tasks to a decision-making module; Achieved improved generalization of lower level learned model

- **Microsoft Research Asia** Jul. 2019 – Oct. 2019
  Research Intern, Advisor: Qiwei Ye *Beijing, China*

  Multi-Agent Reinforcement Learning Game AI for Google Research Football
  - Developed a Multi-agent Game AI for Google Research Football based on QMIX with pretrained model
  - Surveyed and implemented various baseline MARL algorithms such as Independent PPO and QMIX
  - Proposed pre-training agents through imitation learning, which significantly outperformed baselines and led to an improved model that was superior to benchmark

- **Peking University** Jan. 2019 – Apr. 2019
  Undergraduate Research Assistant, Advisor: Zongqing Lu *Beijing, China*

  Teacher-Student Curriculum Learning for Visual Active Tracking
  - Proposed a novel teacher-student curriculum learning approach to train a strong tracker UAV with learning target UAV generating various moving patterns
  - Formulated Multi-UAV tracking as Meta-Learning problem, with the target serving as a meta-agent
  - Designed a RL learning target with two sub policies to generate experienced and novel moving patterns, which resulted in improved robustness of the tracker agent

Selected Open Source Projects

- **Peking University** Apr. 2022 – Oct. 2022
  Research Assistant, Advisor: Yaodong Yang *Beijing, China*

  TorchOpt: An Efficient Library for Differentiable Optimization
  - Developed a differentiable AdamW optimizer for gradient-based Meta-Learning research
  - Implemented MAML examples to verify the effectiveness of differentiable AdamW and compatibility with functorch
  - Designed matrix inversion linear solver with neumann series approximation and implemented implicit MAML omniglot examples with corresponding linear solver
  - Completed a co-first-authored paper published at **JMLR 2023**

- **Sea AI Lab** Apr. 2022 – Oct. 2022
  Outside Collaborator, Advisor: Zhongwen Xu *Beijing, China*

  EnvPool: A Highly Parallel Reinforcement Learning Environment Execution Engine
  - Achieved 3M FPS throughput with MuJoCo physics engine on a single DGX-A100 machine
  - Demonstrated high-performance RL agent training with Atari games and MuJoCo using EnvPool and RL Games
  - Conducted MuJoCo environment alignment test, contributed bug reports, and performed debugging
  - Completed a co-authored paper published at **NeurIPS 2022 Datasets and Benchmarks**

Honors & Awards

Principal Scholarship For Undergraduate Research Jul. 2018

For outstanding undergraduate researchers in Peking University chosen through comprehensive evaluation

Award For Scientific Research Oct. 2017

For students (8 out of 392 in EECS College, PKU) demonstrating exceptional academic performance

Technical Skills

Programming Languages: Python, C/C++, Bash, Go, Stata

Tools and Frameworks: TensorFlow, PyTorch, Git, LaTeX, Docker, TravisCI, Google Cloud Platform, VS Code, Visual Studio, PyCharm