Shiwei Liu

Date of Birth: 1993.01 | Nationality: Chinese | sliu@tue.ellius.eu | Website | Google Scholar

RESEARCH INTERESTS

My overarching research goal is to advance the frontiers of AI from multiple perspectives, including but not limited to enhancing the capabilities of large foundation models via pre-training and post-training (e.g., RLVR, SFT), improving the accessibility of AI via low-dimensionality, developing a deeper understanding of deep learning, exploring novel architectures, and better data curation.

PROFESSIONAL EXPERIENCE

Group Leader *Sep. 2025 - Present*
Max Planck Institute for Intelligent Systems, Tübingen, Germany

Principal Investigator (PI) *Sep. 2025 - Present*
ELLIS Institute Tübingen, Tübingen, Germany

Royal Society Newton International Fellow *Jan. 2024 - Sep. 2025*
Mathematical Institute, University of Oxford, UK

Junior Research Fellow at Somerville College *Sep. 2024 - Sep. 2025*
University of Oxford, UK

Postdoctoral Fellow *Sep. 2022 - Oct. 2023*
University of Texas at Austin (UT Austin), USA

AWARDS AND HONOURS

- **Rising Star Award**, KAUST, Link *01/2024*
- **Rising Star Award**, Conference on Parsimony and Learning 2024, Link *10/2023*
- **Royal Society Newton International Fellowship Award**, UK, Link *01/2024*
- **Best Paper Award**, Learning on Graphs Conference (LoG 2022) *11/2022*
- **Best Ph.D. Dissertation Award Runner-Up**, Informatics Europe, Link *10/2023*
- **Cum Laude (Distinguished Ph.D. thesis)**, Eindhoven University of Technology, 5% *4/2022*
- **Carnegie Bosch Fellowship**, Carnegie Mellon University & Bosch, Link *04/2023*
- **AAAI-23 Travel Scholarship**, USA *12/2022*
- **Outstanding Intern**, JD Explore Academy, 2% *10/2021*
- **The First Class Scholarship**, Harbin Institute of Technology, China *07/2017*
- **China National Scholarship**, China, 0.2% *10/2016*
- **The First Class Scholarship**, Harbin Institute of Technology, China *07/2016*
- **The Outstanding Graduate**, North University of China, China *07/2015*

SELECTIVE PUBLICATION

Sun, Wenfang, Xinyuan Song, Pengxiang Li, Lu Yin, Yefeng Zheng, and **Shiwei Liu**. *The curse of depth in large language models.* NeurIPS 2025. [GitHub]

Lu Yin, You Wu, Zhenyu Zhang, Cheng-Yu Hsieh, Yaqing Wang, Yiling Jia, Mykola Pechenizkiy, Yi Liang, Zhangyang Wang, **Shiwei Liu**. *Outlier Weighed Layerwise Sparsity (OWL): A Missing Secret Sauce for Pruning LLMs to High Sparsity.* ICML 2024. [GitHub]

Tianjin Huang, Tianlong Chen, Meng Fang, Vlado Menkovski, Jiaxu Zhao, Lu Yin, Yulong Pei, Decebal Constantin Mocanu, Zhangyang Wang, Mykola Pechenizkiy, **Shiwei Liu**. *You Can Have Better Graph Neural Networks by Not Training Weights at All: Finding Untrained Graph Tickets.* Learning on Graphs Conference (LoG), 2022. [Best Paper Award]. [GitHub]

Shiwei Liu, Tianlong Chen, Xiaohan Chen, Xuxi Chen, Qiao Xiao, Boqian Wu, Mykola Pechenizkiy, Decebal Mocanu, and Zhangyang Wang. *More Convnets in the 2020s: Scaling Up Kernels Beyond 51×51 Using Sparsity.* ICLR 2023. [GitHub]

Shiwei Liu, Lu Yin, Decebal Constantin Mocanu, and Mykola Pechenizkiy. *Do We Actually Need Dense Over-Parameterization? In-Time Over-Parameterization in Sparse Training.* The Thirty-eighth International Conference on Machine Learning (ICML), PMLR, 2021. [GitHub]

EDUCATION

Eindhoven University of Technology (TU/e), The Netherlands   *Mar. 2018 - Apr. 2022*
Ph.D., Department of Mathematics and Computer Science
w. Cum Laude/Distinguished Dissertations
Promotors: Prof. Dr. Mykola Pechenizkiy; Dr. Decebal Constantin Mocanu

Harbin Institute of Technology (Shenzhen), China   *Sep. 2015 - Jul. 2017*
M.Eng, School of Mechanical Engineering and Automation

North University of China, China   *Sep. 2011 - Jul. 2015*
B.Eng., Control Engineering

GRANTS (TOTAL: $3,371,067.00)

• **Hector Endowed Fellowship**
Funding Body: Hector Foundation & Max Planck Institute
Value of Award: 1.8M Euro
Duration: Feburary 2025 - September 2031
Role on the Grant: PI

• **Newton International Fellowship Award**
Funding Body: Royal Society & British Academy
Value of Award: 394,851 GBP ($481,746)
Duration: January 2024 - January 2027
Role on the Grant: PI

• **NWO Grants for Computing Time**
Funding Body: The Dutch Research Council (NWO)
Value of Award: $90,431.5
Duration: April 2023 - April 2025
Role on the Grant: PI (with Mykola Pechenizkiy and Lu Yin)

• **IFML Postdoctoral Fellowship**
Funding Body: The NSF AI Institute for Foundations of Machine Learning (IFML)
Value of Award: $160,000
Duration: September 2022 - September 2024
Role on the Grant: PI

• **NWO Grants for Computing Time**
Funding Body: The Dutch Research Council (NWO)
Value of Award: $220,000
Duration: April 2022 - April 2024
Role on the Grant: PI (with Mykola Pechenizkiy and Decebal Constantin Mocanu)

- **Carnegie Bosch Fellowships** (awarded but declined)
Funding Body: CMU and Bosch
Value of Award: $160,000
Duration: 2023 - 2025
Role on the Grant: PI

INVITED TALKS

• University of California, Berkeley. Organizer: Michael Mahoney	Jan. 2026
• University of Basel. Organizer: Ivan Dokmanić	Jan. 2026
• University of Tübingen. Organizer: Andreas Geiger	Nov. 2025
• The Chinese University of Hong Kong Shenzhen. Organizer: Ruoyu Sun	Sep. 2025
• Tsinghua University Shenzhen. Organizer: Shutao Xia	Sep. 2025
• ECFI-GAIL International Workshop on Synthetic Data. University of Edinburgh.	June. 2025
• GOSIM AI Paris 2025.	May. 2025
• Embedded Learning and Sensing Systems Group at TU Graz Organizer: Olga Saukh	Mar. 2025
• Computer Science Research Centre. University of Surrey. Organizer: Lu Yin	Jan. 2025
• Optimization Decision Intelligence (ODI) Group. ETH. Organizer: Niao He	Dec. 2024
• Computational Mathematics and Applications Seminar, Math Institute at Oxford.	Nov. 2024
• University of Luxembourg. Organizer: Decebal Constantin Mocanu.	Jun. 2024
• University of Sheffield. Organizer: Nikolaos Aletras, Cass Zhixue Zhao	Jun. 2024
• LTL seminars. University of Cambridge. Organizer: Anna Korhonen, Nigel Collier, and Ivan Vulić	Jun. 2024
• CIMDA-Oxford seminars. University of Oxford. Organizer: Coralia Cartis, Terry Lyons	Apr. 2024
• Data Science Seminar. University of Oxford. Organizer: Jared Tanner	Feb. 2024
• Rising Star in AI Symposium. KAUST. Organizer: Juergen Schmidhuber	Feb. 2024
• LIRA Seminar. Lancaster University. Organizer: Plamen Angelov	Feb. 2024
• Chinese University of Hong Kong-Shenzhen. Organizer: Ruoyu Sun	Jan. 2024
• Rising Star in Cpal Conference. The University of Hong Kong. Organizer: Yi Ma	Jan. 2024
• University of Edinburgh. Organizer: Sotirios Tsaftaris	Jul. 2023
• Zhidx. Organizer: Zhidx	Jul. 2023
• CTSTA Workshop in PLDI 2023. Organizer: Fredrik Kjolstad, Saman Amarasinghe, Michelle Mills Strout	Jun. 2023
• EfficientML Reading Group. Organizer: Olga Saukh	May. 2023
• Sparsity Reading Group. Organizer: Anna Golubeva , Dan Alistarh, Decebal Constantin Mocanu, Gintare Karolina Dziugaite, Utku Evci, Yani Ioannou	Mar. 2023
• Sharc Group, Organizer: Cong Hao	Mar. 2023
• IFML, Organizer: Adam Klivans, Alexandros Dimakis	Jan. 2023
• LoG2022 Meetup, Organizer: Jure Leskovec, Kexin Huang	Dec. 2022
• NCSU Reliable & Efficient Computing Lab, Organizer: Dongkuan Xu	Nov. 2022
• EIC Lab, Organizer: Yingyan Lin	Sep. 2022
• AI XinQingNian	Nov. 2022
• Data Mining Group, Organizer: Mykola Pechenizkiy	May. 2022

EDUCATIONAL AND TEACHING ACTIVITIES

Teaching

• Elements of AI, University of Luxembourg,	Mar. 2025
• 2IMM00 Seminar Data Mining and AI for MSc, Eindhoven University of Technology	Oct. 2023

Tutorial Presentation

- CPAL 2025, "Sparsity and Mixture-of-Experts in the Era of LLMs: A New Odyssey", Stanford, USA, Mar. 2025
- ICASSP 2024, "Sparsity in Large Language Models: The New Odyssey", Seoul, Korea,   Apr. 2024

Co-organizer & Presenter; 75-min tutorial presentation on sparsity in Large Language Models (LLMs).
- VITA group workshop, "Sparsity and Efficiency", Austin, USA   Sep. 2023
Presenter; 60-min tutorial presentation on recent works on sparse and efficient LLMs.
- IJCAI 2023, "Sparse Training for Supervised, Unsupervised, Continual, and Deep Reinforcement Learning with Deep Neural Networks", Macao, China. Link   Aug. 2023
Co-organizer & Presenter; 60-min tutorial presentation from sparse training to sparse scaling.
- ECMLPKDD 2022, "Sparse Neural Network Training", Grenoble, France. Link   Sep. 2022
Co-organizer & Presenter; 45-min tutorial presentation on understanding sparse neural network training with supervised learning.

Students (Co)-Supervising

PhD Students
- Dilxat Muhtar, MPI-IS, ELLIS Institute Tübingen, ETH Zurich
- Wanqi Yang, MPI-IS, ELLIS Institute Tübingen, ETH Zurich
- Alexander Conzelmann, University of Tübingen
- Lu Yin, Eindhoven University of Technology
- Tianjin Huang, Eindhoven University of Technology
- Boqian Wu, University of Twente
- Qiao Xiao, Eindhoven University of Technology
- Ajay Kumar Jaiswal, University of Texas at Austin
- Duc N.M Hoang, University of Texas at Austin
- Zhenyu Zhang, University of Texas at Austin

Master Students - Research Advisor:
- Shu Wang, University of Oxford
- David González Martínez, University of Tübingen
- Keyu Wang, University of Tübingen

Undergraduate Students - Research Advisor:
- Honors Academy: Simon Sukup, Austin Roose, Angelos Mangos, Mikołaj Pujanek, Eleftheria Kolokytha, Eindhoven University of Technology

SERVICE

2026: Program Chair: CPAL 2026
Area Chair: ICLR 2026 Emergency Area Chair, ICML 2026, ICASSP 2026, ARR
2025: Publicity Chair: CPAL 2025
Area Chair: ARR
2024: Area Chair: ICASSP
Conference Reviewer: ICML, ICLR, NeurIPS, AAAI, AISTATS, CPAL
Journal Reviewer: Journal of Selected Topics in Signal Processing, TMLR, TPAMI
2023: Area Chair: ICIP
Conference Reviewer: NeurIPS, ICML, ICLR, CVPR, ICCV, AAAI, UAI, DAC
Journal Reviewer: JMLR, TPAMI, IJCV
2022: Area Chair: ICIP
Conference Reviewer: NeurIPS, ICLR, ICML, CVPR, AAAI
Journal Reviewer: TPAMI
2021: Conference Reviewer: NeurIPS, ICLR, ICML, AISTATS
Journal Reviewer: IEEE Transactions on Evolutionary Computation
2020: Conference Reviewer: ECMLPKDD, IDA
Journal Reviewer: ACM Transactions On Intelligent Systems And Technology

ORGANIZATIONAL CONTRIBUTION

Conference Organization
- Conference on Parsimony and Learning (CPAL), **Program Chair**, Tübingen, Germany. Mar. 2026

- Conference on Parsimony and Learning (CPAL), Publicity Chair, Stanford, USA.   Mar. 2025
- International Conference on Machine Learning and Applications (ICMLA), Special Sessions Chair, Dec. 2024

Workshop Organization

- ICLR 2025, "Workshop on Sparsity in LLMs (SLLM): Deep Dive into Mixture of Experts, Quantization, Hardware, and Inference", Singapore.   May. 2025
- ICLR 2025, "Scalable Optimization for Efficient and Adaptive Foundation Models", Singapore. May. 2025

Competition Organization

- NeurIPS 2024, "Edge-Device Large Language Model Competition", Vancouver, Canada. May. 2024

Tutorial Organization

- CPAL 2025, "Sparsity and MoE in Large Language Models: The New Odyssey", Stanford, USA. March. 2025
- ICASSP 2024, "Sparsity in Large Language Models: The New Odyssey", Seoul, Korea.   Apr. 2024
- IJCAI 2023, "Sparse Training for Supervised, Unsupervised, Continual, and Deep Reinforcement Learning with Deep Neural Networks", Macao, China. Link   Aug. 2023
- ECMLPKDD 2022, "Sparse Neural Network Training", Grenoble, France. Link   Sep. 2022

Meetup Organization

- NeurIPS 2023 Sparsity in Deep Neural Networks Meetup, New Orleans, US.   Dec. 2023
- NeurIPS 2023 Local Meetup, Eindhoven University of Technology, the Netherlands.   Dec. 2023
- ICLR 2022 Local Meetup, Eindhoven University of Technology, the Netherlands.   Apr. 2022

Panellist

- ICASSP 2024, "PROGRESS Workshop", Seoul, Korea.   Apr. 2024
- ICASSP 2024, "Deep Neural Network Model Compression Workshop", Seoul, Korea.   Apr. 2024

MEDIA COVERAGE

- Press Release of Early Career Research Schemes of Royal Society in Establishing the Next Generation of Research Leaders in the UK. Royal Society. Oct. 2023. Link
- Best Dissertation Award Runner-up Award. Informatics Europe. Oct. 2023. Link
- CPAL Rising Star Award. Conference on Parsimony and Learning (CPAL). Oct. 2023. Link
- Shiwei Liu and VITA Group Receive Best Paper Award at LoG 2022. IFML News. Jan. 2023. Link

PEER REVIEWED PUBLICATIONS

Publication Summary

Overall: **78** papers (44 **A*** and 5 **A** conference paper, CORE Conference Ranking), **7** journal papers; among them **19** first-author papers and **17** last-author papers.
Impact: **3039** citations, h-index: **29**, i10-index: **53** (as of 03-Dec-2025).

Peer-Reviewed Conference Publications (reverse chronological order)

[1] Sun, Wenfang, Xinyuan Song, Pengxiang Li, Lu Yin, Yefeng Zheng, and **Shiwei Liu**. *The curse of depth in large language models.* NeurIPS 2025.

[2] He, Di, Ajay Jaiswal, Songjun Tu, Li Shen, Ganzhao Yuan, **Shiwei Liu**, and Lu Yin. *AlphaDecay: Module-wise Weight Decay for Heavy-Tailed Balancing in LLMs.* NeurIPS 2025.

[3] Chen, Tianhao, Xin Xu, Zijing Liu, Pengxiang Li, Xinyuan Song, Ajay Kumar Jaiswal, **Shiwei Liu**, Yu Li, Lu Yin, Can Yang. *GPAS: Accelerating Convergence of LLM Pretraining via Gradient-Preserving Activation Scaling.* NeurIPS 2025.

[4] Pengxiang Li, Lu Yin, Xiaowei Gao, **Shiwei Liu**. *OwLore: Outlier-weighed Layerwise Sampled Low-Rank Projection for Memory-Efficient LLM Fine-tuning.* ACL Findings 2025.

[5] Liu, Zihang, Tianyu Pang, Oleg Balabanov, Chaoqun Yang, Tianjin Huang, Lu Yin, Yaoqing Yang, and **Shiwei Liu**. *LIFT the Veil for the Truth: Principal Weights Emerge after Rank Reduction for Reasoning-Focused Supervised Fine-Tuning.* ICML 2025.

[6] Ajay Jaiswal, Lu Yin, Zhenyu Zhang, **Shiwei Liu**, Jiawei Zhao, Yuandong Tian, Zhangyang Wang. *From GaLore to WeLore: How Low-Rank Weights Non-uniformly Emerge from Low-Rank Gradients.* ICML 2025.

[7] Zhuang, Xialie, Zhikai Jia, Jianjin Li, Zhenyu Zhang, Li Shen, Zheng Cao, and **Shiwei Liu**. *Mask-Enhanced Autoregressive Prediction: Pay Less Attention to Learn More.* ICML 2025.

[8] Zhenyu Zhang, Ajay Jaiswal, Lu Yin, **Shiwei Liu**, Jiawei Zhao, Yuandong Tian, Zhangyang Wang. *Q-GaLore: Quantized GaLore with INT4 Projection and Layer-Adaptive Low-Rank Gradients.* CPAL 2025.

[9] Huang, Tianjin, Ziquan Zhu, Gaojie Jin, Lu Liu, Zhangyang Wang, and **Shiwei Liu**. *SPAM: Spike-Aware Adam with Momentum Reset for Stable LLM Training.* ICLR 2025.

[10] Li, Pengxiang, Lu Yin, and **Shiwei Liu**. *Mix-LN: Unleashing the Power of Deeper Layers by Combining Pre-LN and Post-LN.* ICLR 2025.

[11] Arinbjörn Kolbeinsson, Tianjin Huang, Shanghua Gao, **Shiwei Liu**, Jonathan Richard Schwarz, Anurag Jayant Vaidya, Faisal Mahmood, Marinka Zitnik, Tianlong Chen, Thomas Hartvigsen. *Composable Interventions for Language Models.* ICLR 2025.

[12] Lu, Haiquan*, Yefan Zhou*, **Shiwei Liu**, Zhangyang Wang, Michael W. Mahoney, and Yaoqing Yang. *Alphapruning: Using heavy-tailed self-regularization theory for improved layer-wise pruning of large language models.* NeurIPS 2024.

[13] Zhenyu Zhang, Runjin Chen, **Shiwei Liu**, Zhewei Yao, Olatunji Ruwase, Beidi Chen, Xiaoxia Wu, Zhangyang Wang. *Found in the Middle: How Language Models Use Long Contexts Better via Plug-and-Play Positional Encoding.* NeurIPS 2024.

[14] Boqian Wua, Qiao Xiaob, **Shiwei Liu**, Lu Yin, Mykola Pechenizkiy, Decebal Constantin Mocanu, Maurice Van Keulen, Elena Mocanu. *E2ENet: Dynamic Sparse Feature Fusion for Accurate and Efficient 3D Medical Image Segmentation.* NeurIPS 2024.

[15] Abhinav Bandari, Lu Yin, Cheng-Yu Hsieh, AJAY KUMAR JAISWAL, Tianlong Chen, Li Shen, Ranjay Krishna, **Shiwei Liu**. *Is C4 Dataset Enough for Pruning? An Investigation of Calibration Data for LLM Pruning.* EMNLP 2024

[16] AJAY KUMAR JAISWAL, Bodun Hu, Lu Yin, Yeonju Ro, **Shiwei Liu**, Tianlong Chen, Aditya Akella. *FFN-SkipLLM: A Hidden Gem for Autoregressive Decoding with Adaptive Feed Forward Skipping.* EMNLP 2024

[17] Adriana Fernandez-Lopez, Honglie Chen, Pingchuan Ma, Lu Yin, Qiao Xiao, Stavros Petridis, **Shiwei Liu**, Maja Pantic. *MSRS: Training Multimodal Speech Recognition Models from Scratch with Sparse Mask Optimization.* Interspeech, 2024

[18] Qiao Xiao, Pingchuan Ma, Adriana Fernandez-Lopez, Boqian Wu, Lu Yin, Stavros Petridis, Mykola Pechenizkiy, Maja Pantic, Decebal Constantin Mocanu, **Shiwei Liu**. *Dynamic Data Pruning for Automatic Speech Recognition.* Interspeech, 2024

[19] Lu Yin, You Wu, Zhenyu Zhang, Cheng-Yu Hsieh, Yaqing Wang, Yiling Jia, Mykola Pechenizkiy, Yi Liang, Zhangyang Wang, **Shiwei Liu.** *Outlier Weighed Layerwise Sparsity (OWL): A Missing Secret Sauce for Pruning LLMs to High Sparsity.* The Forty-first International Conference on Machine Learning (ICML), 2024

[20] Lu Yin, Ajay Jaiswal, **Shiwei Liu**, Souvik Kundu, and Zhangyang Wang. *Pruning Small Pre-Trained Weights Irreversibly and Monotonically Impairs "Difficult" Downstream Tasks in LLMs.* The Forty-first International Conference on Machine Learning (ICML), 2024

[21] Yuxin Zhang, Yuxuan Du, Gen Luo, Yunshan Zhong, Zhenyu Zhang, **Shiwei Liu**, Rongrong Ji. *CaM: Cache Merging for Memory-efficient LLMs Inference.* The Forty-first International Conference on Machine Learning (ICML), 2024

[22] Jie Ji, Gen Li, Lu Yin, Minghai Qin, Geng Yuan, Linke Guo, **Shiwei Liu**, Xiaolong Ma. *Advancing Dynamic Sparse Training by Exploring Optimization Opportunities.* The Forty-first International Conference on Machine Learning (ICML), 2024

[23] Zhangheng Li, **Shiwei Liu**, Tianlong Chen, Ajay Kumar Jaiswal, Zhenyu Zhang, Dilin Wang, Raghuraman Krishnamoorthi, Shiyu Chang, Zhangyang Wang. *Sparse Cocktail: Co-Training Many Sparsity Patterns and Ratios at Once.* The Forty-first International Conference on Machine Learning (ICML), 2024

[24] Zhenyu Zhang*, **Shiwei Liu***, Runjin Chen, Bhavya Kailkhura, Beidi Chen, Zhangyang Wang. Q-Hitter: A Better Token Oracle for Efficient LLM Inference via Sparse-Quantized KV Cache. Seventh Conference on Machine Learning and Systems (MLSys), 2024.

[25] Yuxin Zhang, Lirui Zhao, Mingbao Lin, Sun Yunyun, Yiwu Yao, Xingjia Han, Jared Tanner, **Shiwei Liu**, Rongrong Ji. Dynamic Sparse No Training: Training-Free Fine-tuning for Sparse LLMs. The Twelfth International Conference on Learning Representations (ICLR), 2024.

[26] Enneng Yang, Zhenyi Wang, Li Shen, **Shiwei Liu**, Guibing Guo, Xingwei Wang, Dacheng Tao. AdaMerging: Adaptive Model Merging for Multi-Task Learning. The Twelfth International Conference on Learning Representations (ICLR), 2024.

[27] Gen Li, Lu Yin, Jie Ji, Wei Niu, Minghai Qin, Bin Ren, Linke Guo, **Shiwei Liu**, Xiaolong Ma. NeurRev: Train Better Sparse Neural Network Practically via Neuron Revitalization. The Twelfth International Conference on Learning Representations (ICLR), 2024.

[28] Haoyu Ma, Chengming Zhang, Lizhi Xiang, Xiaolong Ma, Geng Yuan, Wenkai Zhang, **Shiwei Liu**, Tianlong Chen, Dingwen Tao, Yanzhi Wang, Zhangyang Wang, Xiaohui Xie. HRBP: Hardware-friendly Regrouping towards Block-wise Pruning for Sparse Training. Conference on Parsimony and Learning (CPAL), 2024 [**Spotlight**].

[29] Lu Yin, Gen Li, Meng Fang, Li Shen, Tianjin Huang, Zhangyang Wang, Vlado Menkovski, Xiaolong Ma, Mykola Pechenizkiy, **Shiwei Liu**. Dynamic Sparsity Is Channel-Level Sparsity Learner. The Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS), 2023. & SNN Workshop at ICLR 2023 [**Spotlight**].

[30] Jaiswal, Ajay, **Shiwei Liu**, Tianlong Chen, and Zhangyang Wang. *The Emergence of Essential Sparsity in Large Pre-trained Models: The Weights that Matter.* The Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS), 2023.

[31] Duc N.M Hoang, Souvik Kundu, **Shiwei Liu**, Zhangyang Wang. Don't just prune by magnitude! Your mask topology is a secret weapon. The Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS), 2023.

[32] Hoang Pham, The-Anh Ta, **Shiwei Liu**, Lichuan Xiang, Dung D. Le, Hongkai Wen, Long Tran-Thanh. *Towards Data-Agnostic Pruning At Initialization: What Makes a Good Sparse Mask?* The Thirty-seventh Conference on Neural Information Processing Systems (NeurIPS), 2023.

[33] Enneng Yang, Li Shen, Zhenyi Wang, **Shiwei Liu**, Guibing Guo, Xingwei Wang. *Data Augmented Flatness-aware Gradient Projection for Continual Learning.* International Conference on Computer Vision (ICCV), 2023.

[34] Ruisi Cai, Xiaohan Chen, **Shiwei Liu**, Jayanth Srinivasa, Myungjin Lee, Ramana Rao Kompella, Zhangyang Wang. *Scaling Federated Learning under Data and Task Heterogeneity: A Pilot Study.* International Conference on Computer Vision (ICCV), MTLFL FedVision Workshop 2023.

[35] Tianjin Huang*, **Shiwei Liu***, Tianlong Chen, Meng Fang, Li Shen, Vlado Menkovski, Lu Yin, Yulong Pei, Mykola Pechenizkiy. *Enhancing Adversarial Training via Reweighting Optimization Trajectory.* The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), 2023.

[36] Jiaxu Zhao, Lu Yin, **Shiwei Liu**, Meng Fang, Mykola Pechenizkiy. *REST: Debiasing Deep Neural Networks through Reweighted Sparse Training.* The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), 2023.

[37] Ajay Kumar Jaiswal, **Shiwei Liu**, Tianlong Chen, Ying Ding, Zhangyang Wang. *Instant Soup: Cheap Pruning Ensembles in A Single Pass Can Draw Lottery Tickets from Large Models.* The Fortieth International Conference on Machine Learning (ICML), PMLR, 2023. [Oral].

[38] Tianjin Huang, Lu Yin, Zhenyu Zhang, Li Shen, Meng Fang, Mykola Pechenizkiy, Zhangyang Wang, **Shiwei Liu**. *Are Large Kernels Better Teachers than Transformers for ConvNets?.* The Fortieth International Conference on Machine Learning (ICML), PMLR, 2023.

[39] Ajay Kumar Jaiswal, **Shiwei Liu**, Tianlong Chen, Ying Ding, Zhangyang Wang. *Graph Ladling: Embarrassingly Scalable and Efficient Training of Powerful GNNs via Data-Centric Model Averaging.* The Fortieth International Conference on Machine Learning (ICML), PMLR, 2023.

[40] **Shiwei Liu**, Tianlong Chen, Xiaohan Chen, Xuxi Chen, Qiao Xiao, Boqian Wu, Mykola Pechenizkiy, Decebal Mocanu, and Zhangyang Wang. *More Convnets in the 2020s: Scaling Up Kernels Beyond 51×51 Using Sparsity.* International Conference on Learning Representations (ICLR), 2023.

[41] Duc N.M Hoang, **Shiwei Liu**, Radu Marculescu, Zhangyang Wang. *Revisiting Pruning at Initialization Through the Lens of Ramanujan Graph.* International Conference on Learning Representations (ICLR), 2023. [Oral].

[42] **Shiwei Liu***, Tianlong Chen*, Zhenyu Zhang, Xuxi Chen, Tianjin Huang, Ajay Kumar Jaiswal, and Zhangyang Wang. *Sparsity May Cry: Let Us Fail (Current) Sparse Neural Networks Together!.* International Conference on Learning Representations (ICLR), 2023. [Spotlight].

[43] Tianlong Chen, Zhenyu Zhang, Ajay Kumar Jaiswal, **Shiwei Liu**, Zhangyang Wang *Sparse MoE with Random Routing as the New Dropout: Training Bigger and Self-Scalable Models.* International Conference on Learning Representations (ICLR), 2023. [Spotlight].

[44] **Shiwei Liu**, and Zhangyang Wang., 2023. *Ten Lessons We Have Learned in the New "Sparseland": A Short Handbook for Sparse Neural Network Researchers.* SNN Workshop at ICLR 2023. [Spotlight].

[45] Lu Yin*, **Shiwei Liu***, Fang Meng, Tianjin Huang, Vlado Menkovski, Mykola Pechenizkiy. *Lottery Pools: Winning More by Interpolating Tickets without Increasing Training or Inference Cost.* Thirty-Sixth AAAI Conference on Artificial Intelligence (AAAI), 2023.

[46] Tianjin Huang, Tianlong Chen, Meng Fang, Vlado Menkovski, Jiaxu Zhao, Lu Yin, Yulong Pei, Decebal Constantin Mocanu, Zhangyang Wang, Mykola Pechenizkiy, **Shiwei Liu**. *You Can Have Better Graph Neural Networks by Not Training Weights at All: Finding Untrained Graph Tickets.* Learning on Graphs Conference (LoG), 2022. [Oral & Best Paper Award].

[47] Qiao Xiao, Boqian Wu, Yu Zhang, **Shiwei Liu**, Mykola Pechenizkiy, Elena Mocanu, Decebal Constantin Mocanu. *Dynamic Sparse Network for Time Series Classification: Learning What to "See".* The Thirty-sixth Conference on Neural Information Processing Systems (NeurIPS), 2022.

[48] Lu Yin, Vlado Menkovski, Meng Fang, Tianjin Huang, Yulong Pei, Mykola Pechenizkiy, **Shiwei Liu**. *Superposing Many Tickets into One: A Performance Booster for Sparse Neural Network Training.* The 38th Conference on Uncertainty in Artificial Intelligence (UAI), 2022.

[49] **Shiwei Liu**, Tianlong Chen, Xiaohan Chen, Li Shen, Decebal Constantin Mocanu, Zhangyang Wang, and Mykola Pechenizkiy. *The Unreasonable Effectiveness of Random Pruning: Return of the Most Naive Baseline for Sparse Training.* International Conference on Learning Representations (ICLR), 2022.

[50] **Shiwei Liu**, Tianlong Chen, Zahra Atashgahi, Xiaohan Chen, Ghada Sokar, Elena Mocanu, Mykola Pechenizkiy, Zhangyang Wang, and Decebal Constantin Mocanu. *Deep Ensembling with No Overhead for either Training or Testing: The All-Round Blessings of Dynamic Sparsity.* International Conference on Learning Representations (ICLR), 2022.

[51] **Shiwei Liu**, Tianlong Chen, Xiaohan Chen, Zahra Atashgahi, Lu Yin, Huanyu Kou, Li Shen, Mykola Pechenizkiy, Zhangyang Wang, and Decebal Constantin Mocanu. *Sparse Training via Boosting Pruning Plasticity with Neuroregeneration.* The Thirty-fifth Conference on Neural Information Processing Systems (NeurIPS), 2021.

[52] **Shiwei Liu**, Lu Yin, Decebal Constantin Mocanu, and Mykola Pechenizkiy. *Do We Actually Need Dense Over-Parameterization? In-Time Over-Parameterization in Sparse Training.* The Thirty-eighth International Conference on Machine Learning (ICML), PMLR, 2021.

[53] **Shiwei Liu**, Decebal Constantin Mocanu, Yulong Pei, and Mykola Pechenizkiy. *Selfish sparse RNN training.* The Thirty-eighth International Conference on Machine Learning (ICML), PMLR, 2021.

[54] **Shiwei Liu**, Tim Van der Lee, Anil Yaman, Zahra Atashgahi, Davide Ferraro, Ghada Sokar, Mykola Pechenizkiy, and Decebal Constantin Mocanu. *Topological Insights into Sparse Neural Networks.* The European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML-PKDD), Ghent, Belgium, 2020.

[55] **Shiwei Liu**. *Learning Sparse Neural Networks for Better Generalization.* 29th International International Joint Conference on Artificial Intelligence (IJCAI), 2020. Doctoral Consortium.

[56] **Shiwei Liu**, Decebal Constantin Mocanu, and Mykola Pechenizkiy. *On improving deep learning generalization with adaptive sparse connectivity.* The Workshop of Understanding and Improving Generalization in Deep Learning, The Thirty-sixth International Conference on Machine Learning (ICML), PMLR, 2019.

[57] Yan, Fulong, **Shiwei Liu**, and Nicola Calabretta. *Network Performance Optimization with Real Time Traffic Prediction in Data Center Network.* 2020 European Conference on Optical Communications (ECOC). IEEE, 2020.

Peer-Reviewed Journal Publications

[58] Yang, Enneng, Li Shen, Zhenyi Wang, **Shiwei Liu**, Guibing Guo, Xingwei Wang, and Dacheng Tao. 2025. Revisiting Flatness-aware Optimization in Continual Learning with Orthogonal Gradient Projection. IEEE Transactions on Pattern Analysis and Machine Intelligence.

[59] **Shiwei Liu**, Yuesong Tian, Tianlong Chen, Li Shen. *Don't Be So Dense: Sparse-to-Sparse GAN Training Without Sacrificing Performance.* International Journal of Computer Vision (IJCV).

[60] Zahra Atashgahi, Xuhao Zhang, Neil Kichler, **Shiwei Liu**, Lu Yin, Mykola Pechenizkiy, Raymond Veldhuis, Decebal Constantin Mocanu. *Feature Selection with Neuron Evolution.*, TMLR. 2023.

[61] Zahra Atashgahi, Joost Pieterse, **Shiwei Liu**, Decebal Constantin Mocanu, Raymond Veldhuis, Mykola Pechenizkiy. *A Brain-inspired Algorithm for Training Highly Sparse Neural Networks.* Machine Learning Journal (ECML-PKDD 2022 journal track).

[62] **Shiwei Liu**, Decebal Constantin Mocanu, Amarsagar Reddy Ramapuram Matavalam, Yulong Pei, and Mykola Pechenizkiy. (2021). *Sparse evolutionary deep learning with over one million artificial neurons on commodity hardware.* Neural Computing and Applications, 1-16.

[63] **Shiwei Liu**, Iftitahu Ni'mah, Vlado Menkovski, Decebal Constantin Mocanu, and Mykola Pechenizkiy. (2021). *Efficient and effective training of sparse recurrent neural networks.* Neural Computing and Applications, 1-12.

Pre-print and Under-Review

[64] Li, Pengxiang, Yefan Zhou, Dilxat Muhtar, Lu Yin, Shilin Yan, Li Shen, Yi Liang, Soroush Vosoughi, and **Shiwei Liu**. *Diffusion language models know the answer before decoding.* arXiv preprint arXiv:2508.19982 (2025).

[65] Wang, Keyu, Tian Lyu, Guinan Su, Jonas Geiping, Lu Yin, Marco Canini, and **Shiwei Liu**. *When Fewer Layers Break More Chains: Layer Pruning Harms Test-Time Scaling in LLMs.* arXiv preprint arXiv:2510.22228 (2025).

[66] Xu, Xin, Tianhao Chen, Fan Zhang, Wanlong Liu, Pengxiang Li, Ajay Kumar Jaiswal, Yuchen Yan et al. *Double-Checker: Enhancing Reasoning of Slow-Thinking LLMs via Self-Critical Fine-Tuning.* arXiv preprint arXiv:2506.21285 (2025).

[67] Su, Guinan, Li Shen, Lu Yin, **Shiwei Liu**, Yanwu Yang, and Jonas Geiping. *GPTailor: Large Language Model Pruning Through Layer Cutting and Stitching.* arXiv preprint arXiv:2506.20480 (2025).

[68] Wang, Zihan, Rui Pan, Jiarui Yao, Robert Csordas, Linjie Li, Lu Yin, Jiajun Wu, Tong Zhang, Manling Li, and **Shiwei Liu**. *Chain-of-Experts: Unlocking the Communication Power of Mixture-of-Experts Models.* arXiv preprint arXiv:2506.18945 (2025).

[69] Xiao, Qiao, Alan Ansell, Boqian Wu, Lu Yin, Mykola Pechenizkiy, **Shiwei Liu**, and Decebal Constantin Mocanu. *Leave it to the Specialist: Repair Sparse LLMs with Sparse Fine-Tuning via Sparsity Evolution.* arXiv preprint arXiv:2505.24037 (2025).

[70] Li, Kechen, Wenqi Zhu, Coralia Cartis, Tianbo Ji, and **Shiwei Liu**. *Sos1: O1 and r1-like reasoning llms are sum-of-square solvers.* arXiv preprint arXiv:2502.20545 (2025).

[71] Luo, Haotian, Li Shen, Haiying He, Yibo Wang, **Shiwei Liu**, Wei Li, Naiqiang Tan, Xiaochun Cao, and Dacheng Tao. *O1-Pruner: Length-Harmonizing Fine-Tuning for O1-Like Reasoning Pruning." arXiv preprint arXiv:2501.12570 (2025). Under Review.*

[72] Huang, Tianjin, Haotian Hu, Zhenyu Zhang, Gaojie Jin, Xiang Li, Li Shen, Tianlong Chen, Lu Liu, Qingsong Wen, Zhangyang Wang, **Shiwei Liu**. *Stable-SPAM: How to Train in 4-Bit More Stably than 16-Bit Adam.* Under Review.

[73] Huang, Tianjin and Meng, Fang and Shen, Li and Liu, Fan and Pei, Yulong and Pechenizkiy, Mykola and **Shiwei Liu** and Chen, Tianlong. *(PASS) Visual Prompt Locates Good Structure Sparsity through a Recurrent HyperNetwork.* Under Review.

[74] Anke Tang, Li Shen, Yong Luo, **Shiwei Liu**, Han Hu, Bo Du, Dacheng Tao. *Data-Adaptive Weight-Ensembling for Multi-Task Model Fusion.* Under Review.

[75] Tianjin Huang, Tianlong Chen, Zhangyang Wang, **Shiwei Liu**. *The Counterattack of CNNs in Self-Supervised Learning: Larger Kernel Size Might be All You Need.* Under Review.

[76] Tiansheng Huang, **Shiwei Liu**, Li Shen, Fengxiang He, Weiwei Lin, Dacheng Tao., 2022. *Achieving Personalized Federated Learning with Sparse Local Models. arXiv preprint arXiv:2201.11380.*

[77] Yin Nan, Li Shen, Mengzhu Wang, **Shiwei Liu**, Chong Chen, Xian-Sheng Hua, Xiao Luo. *SPORT: A Subgraph Perspective on Graph Classification with Label Noise.* Under Review.

Ph.D. Thesis

[78] **Shiwei Liu**. Sparse Neural Network Training with In-Time Over-Parameterization. The Eindhoven University of Technology. 2022. Paper