Haoyu He

Autonomous Vision Group
University of Tübingen
Tübingen AI Center, Germany, 72076

haoyu.he@uni-tuebingen.de
https://cli212.github.io/
Google Scholar

Education

University of Tübingen, Tübingen, Germany *2022 - present*
PhD student in Machine Learning
Supervisor: Prof. Andreas Geiger
Research Directions: diffusion language models; efficient architectures for long sequence modeling; self-play agents for scientific discovery.

Northeastern University, Boston, USA *2020 - 2022*
Msc in Artificial Intelligence
Supervisors: Prof. Cheng Tan & Prof. Raman Chandrasekar
Research Experience: Worked on text simplification and neural network verification for systems, specifically.

Wuhan University of Science and Technology, Wuhan, China *2015 - 2019*
Bsc in Computer Science and Technology

Selected Publications

4. **MDPO: Overcoming the Training-Inference Divide of Masked Diffusion Language Models**.
Haoyu He, Katrin Renz, Yong Cao, Andreas Geiger
We introduce a policy optimization method for post-training masked diffusion language models to mitigate the pretraining-inference gap. In addition, we design a training-free remasking strategy for flexible generation. *arXiv, 2025.*

3. **NN4SysBench: Characterizing Neural Network Verification for Computer Systems**.
Shuyi Lin, **Haoyu He**, Tianhao Wei, Kaidi Xu, Huan Zhang, Gagandeep Singh, Changliu Liu, Cheng Tan
*Neural Information Processing Systems (**NeurIPS**), 2024.*

2. **HDT: Hierarchical Document Transformer.**
Haoyu He, Markus Flicke, Jan Buchmann, Iryna Gurevych, Andreas Geiger
*Conference on Language Modeling (**COLM**), 2024.*
A sparse Transformer that exploits hierarchical document structure via multi-level sparse attention, including a custom Triton kernel for efficient sparse attention computation, improving sample efficiency and downstream task performance.

1. **Distiller: A Systematic Study of Model Distillation Methods in Natural Language Processing**.
Haoyu He, Xingjian Shi, Jonas Mueller, Sheng Zha, Mu Li, George Karypis
*The Second Workshop on Simple and Efficient Natural Language Processing @ **EMNLP**, 2021.*

Work Experience

Amazon AI Lab, Shanghai China *2020 - 2021*
Research Intern

- Introduced a meta-learning framework for knowledge distillation (KD) of Transformer-based language models. Unified existing KD objectives as instances of maximizing bounds of the mutual information (MI) and proposed a novel objective function to boost knowledge transfer based on MI estimation. Additionally, we designed the first automated KD algorithm *AutoDistiller* that automatically predict a good KD pipeline given a new dataset.
- Aforementioned work is referenced by Stanford CS329p, 2021 fall and accepted at SustaiNLP 2021, EMNLP.

Supervisors: Dr. Xingjian Shi & Sheng Zha

Software

1. **Scholar Inbox**
Markus Flicke, Glenn Angrabeit, Madhav Iyengar, Vitalii Protsenko, Illia Shakun, Jovan Cicvaric, Bora Kargi, **Haoyu He**, Lukas Schuler, Lewin Scholz, Kavyanjali Agnihotri, Yong Cao, Andreas Geiger
Desc: A personal paper recommender which enables researchers to stay up-to-date with the most relevant progress in their field based on their personal research interests. We have supported ML conferences such as CVPR, ECCV, ICML, ICLR, and ACL with the conference planner feature to assist participants find relevant posters.
Stats as of Feb 2026: >37,000.

Selected Awards and Honors

- Northeastern University, Khoury Research Scholarship *2021-2022*