Tianxing He

Address: 185 E Stevens Way NE, Seattle, WA 98195-2350, United States
Email: goosehe@cs.washington.edu
Homepage: https://cloudygoose.github.io
Google-Scholar-Link: https://scholar.google.com/citations?hl=en&user=egmfjjwAAAAJ

EDUCATION & RESEARCH INTERESTS

University of Washington, WA, USA.
Postdoc, Computer Science. July 2022 – Current

- Advisor: Prof. Yulia Tsvetkov.

Massachusetts Institute of Technology, MA, USA.
Ph.D., Computer Science. Sep 2017 – May 2022

- Advisor: Prof. James Glass.
- Thesis: *Towards a Deeper Understanding of Neural Language Generation*.

Shanghai Jiao Tong University, Shanghai, China.
M.S., Computer Science. Sept 2014 – June 2017

- Advisor: Prof. Kai Yu.
- Dissertation: *Structured RNNLM and its Applications in Automatic Speech Recognition*.

B.S., Computer Science (ACM Honors Class). Sept 2010 – June 2014

Research Interests: Trustworthy Generative AI, (Large) Language Models.

PUBLICATIONS (* means co-first-author)

SemStamp: A Semantic Watermark with Paraphrastic Robustness for Text Generation.
Tianxing He*, Abe Bohan Hou*, Jingyu Zhang*, Yichen Wang, Yung-Sung Chuang, Hongwei Wang, Lingfeng Shen, Benjamin Van Durme, Daniel Khashabi, and Yulia Tsvetkov.
The 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL 2024**).

LatticeGen: A Cooperative Framework which Hides Generated Text in a Lattice for Privacy-Aware Generation on Cloud.
Tianxing He*, Mengke Zhang*, Tianle Wang, Lu Mi, Fatemehsadat Mireshghallah, Binyi Chen, Hao Wang, and Yulia Tsvetkov.
Findings of the 2024 Annual Conference of the North American Chapter of the Association for Computational Linguistics (**NAACL-Findings 2024**).

On the Blind Spots of Model-Based Evaluation Metrics for Text Generation.
Tianxing He*, Jingyu Zhang*, Tianle Wang, Sachin Kumar, Kyunghyun Cho, James Glass, and Yulia Tsvetkov.
The 61th Annual Meeting of the Association for Computational Linguistics (**ACL 2023**).

Knowledge Card: Filling LLMs' Knowledge Gaps with Plug-in Specialized Language Models.
Shangbin Feng, Weijia Shi, Yuyang Bai, Vidhisha Balachandran, **Tianxing He**, and Yulia Tsvetkov.
The Twelfth International Conference on Learning Representations (**ICLR 2024**).

On the Zero-Shot Generalization of Machine-Generated Text Detectors.
Xiao Pu, Jingyu Zhang, Xiaochuang Han, Yulia Tsvetkov, and **Tianxing He**.
The 2023 Conference on Empirical Methods in Natural Language Processing (**EMNLP-Findings 2023**).

Can Language Models Solve Graph Problems in Natural Language?

Heng Wang*, Shangbin Feng*, **Tianxing He**, Zhaoxuan Tan, Xiaochuang Han, and Yulia Tsvetkov.
The Thirty-seventh Conference on Neural Information Processing Systems (**NeurIPS 2023**).

Learning Time-Invariant Representations for Individual Neurons from Population Dynamics.
Lu Mi, Trung Le, **Tianxing He**, Eli Shlizerman, and Uygar Sümbül.
The Thirty-seventh Conference on Neural Information Processing Systems (**NeurIPS 2023**).

PCFG-based Natural Language Interface Improves Generalization for Controlled Text Generation.
Jingyu Zhang, James Glass and **Tianxing He**.
The 12th Joint Conference on Lexical and Computational Semantics (**StarSEM 2023**).
Also accepted at the Efficient Natural Language and Speech Processing Workshop (**NeurIPS-ENLSP 2022**).
The Best Paper Award at the Workshop.

Controlling the Focus of Pretrained Language Generation Models.
Jiabao Ji, Yoon Kim, James Glass and **Tianxing He**.
The 60th Annual Meeting of the Association for Computational Linguistics (**ACL-Findings 2022**).

Exposure Bias versus Self-Recovery: Are Distortions Really Incremental for Autoregressive Text Generation?
Tianxing He, Jingzhao Zhang, Zhiming Zhou, and James Glass.
The 2021 Conference on Empirical Methods in Natural Language Processing (**EMNLP 2021**).

Joint Energy-based Model Training for Better Calibrated Natural Language Understanding Models.
Tianxing He, Bryan McCann, Caiming Xiong and Ehsan Hosseini-Asl.
The 16th Conference of the European Chapter of Association for Computational Linguistics (**EACL 2021**).

Analyzing the Forgetting Problem in the Pretrain-Finetuning of Dialogue Response Models.
Tianxing He, Jun Liu, Kyunghyun Cho, Myle Ott, Bing Liu, James Glass and Fuchun Peng.
The 16th Conference of the European Chapter of Association for Computational Linguistics (**EACL 2021**).

A Systematic Characterization of Sampling Algorithms for Open-ended Language Generation.
Tianxing He*, Moin Nadeem*, Kyunghyun Cho and James Glass.
The 1st Conference of the Asia-Pacific Chapter of Association for Computational Linguistics (**AACL 2020**).

Why Gradient Clipping Accelerates Training: A Theoretical Justification for Adaptivity.
Jingzhao Zhang, **Tianxing He**, Suvrit Sra and Ali Jadbabaie.
The Eighth International Conference on Learning Representations (**ICLR 2020**). Reviewer Scores: **8/8/8**.

Negative Training for Neural Dialogue Response Generation.
Tianxing He and James Glass.
The 58th Annual Meeting of the Association for Computational Linguistics (**ACL 2020**).

An Empirical Study of Transformer-based Neural Language Model Adaptation.
Ke Li, Zhe Liu, **Tianxing He**, Hongzhao Huang, Fuchun Peng, Daniel Povey and Sanjeev Khudanpur.
The 45th International Conference on Acoustics, Speech, and Signal Processing (**ICASSP 2020**).

Detecting Egregious Responses in Neural Sequence-to-sequence Models.
Tianxing He and James Glass.
The Seventh International Conference on Learning Representations (**ICLR 2019**).

Multi-View LSTM Language Model with Word-Synchronized Auxiliary Feature for LVCSR.
Yue Wu, **Tianxing He**, Zhehuai Chen, Yanmin Qian, and Kai Yu.
The 2017 Chinese Computational Linguistics and Natural Language Processing (**CCL 2017**).

On Training Bi-directional Neural Network Language Model with Noise Contrastive Estimation.
Tianxing He, Yu Zhang, Jasha Droppo and Kai Yu.
The 10th International Symposium on Chinese Spoken Language Processing (**ISCSLP 2016**).

Exploiting LSTM Structure in Deep Neural Networks for Speech Recognition.
Tianxing He and Jasha Droppo.
The 41st IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP 2016**).

RNN Language Model with Structured Word Embeddings for Speech Recognition.
Tianxing He, Xu Xiang, Yanmin Qian and Kai Yu.
The 2015 IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP 2015**).

Automatic Model Redundancy Reduction for Fast Back-Propagation for Deep Neural Networks in Speech Recognition.
Yanmin Qian, **Tianxing He**, Wei Deng, and Kai Yu.
The 2015 International Joint Conference on Neural Networks (**IJCNN 2015**).

Paragraph Vector Based Topic Model for Language Model Adaptation.
Wengong Jin, **Tianxing He**, Yanmin Qian, and Kai Yu.
The Sixteenth Conference of the International Speech Communication Association (**InterSpeech 2015**).

Reshaping Deep Neural Network for Fast Decoding by Node-Pruning.
Tianxing He, Yuchen Fan, Yanmin Qian, Tian Tan and Kai Yu.
2014 IEEE International Conference on Acoustics, Speech and Signal Processing (**ICASSP 2014**).

WORK UNDER SUBMISSION

Stumbling Blocks: Stress Testing the Robustness of Machine-Generated Text Detectors Under Attacks.
Yichen Wang, Shangbin Feng, Abe Bohan Hou, Xiao Pu, Chao Shen, Xiaoming Liu, Yulia Tsvetkov, and **Tianxing He**.

k-SEMSTAMP : A Clustering-Based Semantic Watermark for Detection of Machine-Generated Text.
Abe Bohan Hou, Jingyu Zhang, Yichen Wang, Daniel Khashabi, and **Tianxing He**.

Resolving Knowledge Conflicts in Large Language Models.
Yike Wang, Shangbin Feng, Heng Wang, Weijia Shi, Vidhisha Balachandran, **Tianxing He**, and Yulia Tsvetkov.

RESEARCH AWARDS, SCHOLARSHIPS AND HONORS

The UW Postdoc Research Award 2023 ($10,000).
CCF-Tencent Rhino-Bird Young Faculty Open Research Fund 2023 (with Prof. Yulia Tsvetkov, $50,000).
The ORACLE Project Award 2023 ($100,000 of cloud credits).
The Best Paper Award at the Efficient Natural Language and Speech Processing Workshop (NeurIPS-ENLSP 2022).
The Ho Ching and Han Ching Scholarship Award (MIT EECS, 2019).
SJTU Outstanding Bachelor Thesis (1% selected in SJTU, 2014).
CCF Outstanding Undergraduates (100 selected in China, 2014).
National Olympics of Information Science, Silver Medal (Top 100 in China, 2010).

TEACHING

Two guest lectures in UW 447+547 (NLP) with title *Neural Network Language Modeling* (2022).
Three Guest lectures to SJTU ACM Class with title *My Story with Language Models* (2021 & 2022 & 2023).
Guest lecture to SJTU ACM Class with title *A Gentle Introduction to Modern NLP* (2020).
Guest lecture in MIT 6.864 with title *Advanced Language Modeling* (2020).
Guest Lecture on *Recurrent Neural Networks* in the Deep Learning Course, for ACM Class (2016).
TA for MIT 6.864 (Advanced Natural Language Processing) (2020).
TA for the Deep Learning Course for SJTU ACM Class (2016).

Four TAs for ACM Class, in courses related to programming, algorithms, (head TA) data structures, and (head TA) compilers (2012-2014).

DIVERSITY, EQUITY, AND INCLUSION (DEI)

Serving as Mentor for Institute for African-American Mentoring in Computing Sciences (IAAMCS, 2023). Around half of my mentored students (listed below) are from under-represented groups such as African-American, first-generation, Asian female, etc.

MENTORED STUDENTS

Wengong Jin (SJTU undergrad) -> MIT (PhD). Yue Wu (SJTU master) -> Alibaba.
Jiabao Ji (SJTU undergrad) -> UCSB (PhD). Moin Nadeem (MIT master) -> MosaicML.
Tianle Wang (UCSD master) -> Applying PhD. Hanwen Shi (SJTU undergrad) -> SJTU (PhD).
Jingyu (Jack) Zhang (JHU undergrad) -> JHU (PhD). Yichen Wang (XJTU undergrad) -> Applying PhD.
Lu Mi (MIT PhD) -> UW (Postdoc). Xiao (Sophia) Pu (PKU undergrad) -> Applying PhD.
Abe Bohan Hou (JHU undergrad) -> Applying PhD.
Vincent Davidson (Clemson PhD) -> Applying Postdoc/faculty.
Charles Bickham (USC PhD) -> Applying Postdoc/faculty.

ACADEMIC SERVICES

Reviewer for ICLR 2024, EMNLP 2023, FAccT 2023, ICML 2023, ACL 2023, EMNLP 2022, ACL Rolling Review 2021, ICLR 2022, AAAI 2022, ICLR 2021, ICML 2021, ACL 2021, EMNLP 2021, Neurips 2021, Neurips 2020.
Reviewer for the 2021 MIT graduate program application.
Chief Editor of SpeechLab (led by Prof. Kai Yu) magazine *Tech reports* (2014-2017).