

---

EDUCATION	<b>University of Science and Technology of China (USTC)</b> <span style="float: right;">Hefei, China</span> <i>B.Sc. in Statistics School of the Gifted Young (special program)</i> Sep. 2021 - Jul. 2025 <ul style="list-style-type: none"><li>Enrolled one year younger than typical students</li></ul>
PUBLICATION	<ol style="list-style-type: none"><li><b>Yi Yang</b>, Yiming Wang, ZhiQiang Tang, Jiahong Yuan. Automated Tone Transcription and Clustering with Tone2Vec. <i>Findings of the Association for Computational Linguistics: EMNLP</i>, 2024.</li><li>Jiancan Wu*, <b>Yi Yang*</b>, Yuchun Qian, Yongduo Sui, Xiang Wang, Xiangnan He. GIF: A General Graph Unlearning Strategy via Influence Function. <i>Proceedings of the ACM Web Conference (WWW)</i>, 2023. <i>Peer-reviewed non-archival presentations</i></li><li><b>Yi Yang</b>, Yiming Wang, Jiahong Yuan. Saving Voices: How AI Can Rescue Endangered Languages in the Digital World? <i>74th Annual International Communication Association Conference, Beijing Regional Hub (ICA Beijing)</i>, 2024.</li><li><b>Yi Yang</b>, Yiming Wang, Jiahong Yuan. Automatic Transcription and Representations for Lexical Tones in Sino-Tibetan Languages. <i>10th International Conference on Computational Social Science (IC2S2)</i>, 2024.</li></ol>
PREPRINT	<ol style="list-style-type: none"><li><b>Yi Yang*</b>, Yiming Wang*, Jiahong Yuan. Transformer-based Speech Model Learns Well as Infants and Encodes Abstractions through Exemplars in the Poverty of the Stimulus Environment. In submission to <i>COLING 2025</i>.</li><li>Yiming Wang, <b>Yi Yang</b>, Jiahong Yuan. Normalization through Fine-tuning: Understanding Wav2vec 2.0 Embeddings for Phonetic Analysis. In submission to <i>ICASSP 2025</i>.</li><li>Zizhao Zhang*, <b>Yi Yang*</b>, Lutong Zou*, He Wen*, Tao Feng, Jiaxuan You. RDBench: ML Benchmark for Relational Databases. <i>arXiv:2310.16837</i>.</li></ol>
RESEARCH EXPERIENCE	<b>University of Pennsylvania, Advisor: Prof. Mark Liberman</b> <span style="float: right;">Jul. 2024 - Sep. 2024</span> <i>Is Bayesian Phylogenetics Really Reliable for Language Evolutions?</i> <i>Keywords: language evolution, bayesian phylogenetics, cognate sets</i> <ul style="list-style-type: none"><li>Generated cognate datasets and proposed quantitative metrics to evaluate Bayesian Phylogenetics with ground truth, revisited the mathematics behind.</li></ul> <b>USTC, Advisor: Prof. Jiahong Yuan</b> <span style="float: right;">Oct. 2023 - Oct. 2024</span> <i>Can Machines Perceive Speech in Human-like Poverty of the Stimulus Environments?</i> <i>Keywords: language acquisition, wav2vec2.0, speech recognition</i> <ul style="list-style-type: none"><li>Designed sparsity and noise scenarios on the phoneme and tone recognition to simulate the Poverty of the Stimulus environments along with three metrics for abstraction: label correction, categorical patterns, and clustering effects</li><li><i>wav2vec2.0</i> can learn, correct, and re-represent exemplars—speech and labels—into abstractions as parameters, moving beyond simple memorization.</li></ul> <i>Automated Tone Toolkit for Low-resource Indigeneous Sino-Tibetan Languages</i> <ul style="list-style-type: none"><li>Proposed the first automated tone transcription and clustering methods for documentation, pitch-based similarity representations Tone2Vec for analysis</li><li>Released ToneLab, to facilitate automated fieldwork and cross-regional analysis.</li><li>Experiments demonstrate that these algorithms are especially beneficial for low resources indigeneous languages, which perform well in transcription and clustering with a small amount of data.</li></ul>

RESEARCH EXPERIENCE	<p>UIUC, Advisor: Prof. Jiaxuan You <span style="float: right;">May. 2023 - Oct. 2023</span>  <i>Machine Learning for Relational Databases</i></p> <ul style="list-style-type: none"> <li>• Defined machine learning tasks as the column value prediction for relational databases and transformed databases into graphs</li> <li>• Collected hierarchical datasets along and designed multiple tasks to enable meaningful comparisons between ML methods from diverse domains</li> </ul>
	<p>USTC, Advisor: Prof. Xiangnan He <span style="float: right;">May. 2022 - May. 2023</span>  <i>Machine Unlearning for Structural Data Privacy</i>  <i>Keywords: responsible AI, machine unlearning, interpretability, graph neural networks</i></p> <ul style="list-style-type: none"> <li>• Presented a unified problem formulation of diverse graph unlearning tasks w.r.t. node, edge, and feature by different privacy and security requests.</li> <li>• Proposed GIF, a model-agnostic unlearning method for graphs, which considered the inter-dependency between connected neighbors.</li> <li>• Deduced the closed-form solution of parameter changes on one-layer graph convolution networks to provide a better understanding of the unlearning mechanism</li> </ul>
AWARDS AND HONORS	<ul style="list-style-type: none"> <li>• <b>Baogang Special Merit Scholarship (15 undergrads for innovation in China) 2024</b></li> <li>• <b>First Prize of the Chinese Mathematics Competitions (top 8%), Anhui division 2022</b></li> <li>• <b>Outstanding Student Scholarship</b> × 2, USTC <span style="float: right;">2022, 2023</span></li> </ul>
SKILLS	<p><b>Languages:</b> Chinese (Standard Mandarin, Jianghuai Mandarin), English  <b>Programming:</b> Python, C++, R, machine learning frameworks like PyTorch, Bayesian phylogenetic analysis software BEAST, familiar with Linux</p>
SERVICES	<p><b>Reviewers for:</b> ICLR 2025, COLING 2025, ICLR 2024 AGI Workshop  Editor of Ling-Lunch, the most well-known linguistics WeChat public account in China.</p>